Unexpected mutual regulation underlies paralogue functional diversification and promotes maturation of a protective epithelial tissue

Daniela Gurska [1], Iris M. Vargas Jentzsch [1], Kristen A. Panfilio [1,2]*

[1] Institute of Zoology: Developmental Biology, University of Cologne, 50674 Cologne, Germany
[2] School of Life Sciences, University of Warwick, Coventry, CV4 7AL, United Kingdom

* Corresponding author: Kristen.Panfilio@alum.swarthmore.edu

ABSTRACT

Insect *Hox3/zen* genes represent an evolutionary hotspot for changes in function and copy number. Single orthologues are required either for early specification or late morphogenesis of the extraembryonic tissues, which protect the embryo. The tandemly duplicated *zen* paralogues of the beetle *Tribolium castaneum* present a unique opportunity to investigate both functions in a single species. We dissect the paralogues' expression dynamics (transcript and protein) and transcriptional targets (RNA-seq after RNAi) throughout embryogenesis. We identify an unexpected role of Tc-Zen2 in repression of *Tc-zen1*, generating a negative feedback loop that promotes developmental progression. Tc-Zen2 regulation is dynamic, including within co-expressed multigene loci. We also show that extraembryonic development is the major event within the transcriptional landscape of late embryogenesis and provide a global molecular characterization of the extraembryonic serosal tissue. Altogether, we propose that paralogue mutual regulation arose progressively and drove multiple instances of *zen* subfunctionalization, leading to complementary extant roles.

Change over macroevolutionary time scales can produce new gene functions, with the *Hox3/zen* genes of insects representing a striking example. Across the bilaterian animals, Hox genes are conserved in genomic organization, expression, and function, with roles in tissue specification along the anterior-posterior body axis of the developing embryo [1]. Instead, the *Hox3* genes in winged insects, known as *zen*, are prone to genomic microinversions [2-4], and they are required in the novel tissue domain of the extraembryonic membranes (EEMs) [5].

The EEMs, simple (monolayer) epithelia, are an evolutionary innovation to protect the developing insect. Initially they surround the early embryo, forming a multilayered barrier from the environment, with an outer serosa and inner amnion [5]. In particular, the serosa is capable of innate immune responses [6,7] and it secretes a thick chitin-based cuticle that mechanically reinforces the eggshell and provides desiccation resistance [8-11]. However, in later development the EEMs must actively withdraw to ensure correct closure of the body [12,13].

Functional studies have identified roles for *zen* in either early EEM specification or late EEM withdrawal (reviewed in [14]). Although the *Hox3* locus is prone to lineage-specific duplications [15,16], to date a single EEM function – specification or morphogenesis (late tissue remodeling for withdrawal) – is known per species in bugs and flies [17-19]. This is even true in the derived case of the fruit fly *Drosophila melanogaster*, which has three functionally distinct paralogues: *zen* itself is involved in EEM specification, the duplicate *z2* is not required for embryogenesis, and the dipteran-specific *bicoid* encodes a maternal determinant with no extraembryonic role [18,20-23]. Furthermore, secondary tissue simplification of the EEMs in *Drosophila* obviated the requirement for the late withdrawal function [14]. Thus, the original role of *zen* within the extraembryonic domain has been obscured by ongoing evolutionary changes in both *zen* and the EEMs.

There is a notable exception to the pattern of a single EEM role of *zen* per species. In the red flour beetle, *Tribolium castaneum*, *zen* has undergone a tandem duplication. *Tc-zen1* was first cloned from cDNA [24], while *Tc-zen2* was later identified by sequencing the Hox cluster directly [25]. The paralogues are striking for their compact, shared gene structure and for their proximity: within the 58-kb region between *Hox2/mxp* and *Hox4/Dfd,* the paralogues occupy a <3-kb interval, with only 216 bp between the 3′ UTR of *Tc-zen1* and the initiation codon of *Tc-zen2* [25]. Nonetheless, subsequent functional diversification has equipped the paralogues with either of the two known EEM functions: early-acting *Tc-zen1* specifies the serosal tissue, while *Tc-zen2* is required for late EEM withdrawal morphogenesis [12,26]. We thus undertook a detailed molecular characterization of the beetle paralogues to elucidate the evolutionary history of *zen* functional changes, and more generally to shed light on the regulatory implications of paralogue retention and diversification after a gene duplication event.

Here we present differences in the regulation of *Tc-zen1* and *Tc-zen2* as well as in their own transcriptional signatures as homeodomain transcription factors, providing the first detailed functional dissection of insect *zen* duplicates outside of *Drosophila*. Surprisingly, peak expression does not coincide with the time of primary function – detectable morphologically and transcriptionally – for *Tc-zen2*, which despite its late role has strong early expression like *Tc-zen1*. Yet, instead of a lack of function or shadow redundancy to Tc-Zen1, we uncover a distinct early role of Tc-Zen2 in the regulation of a key subset of genes, including the rapid repression of *Tc-zen1*. RNA-seq data also reveal subtle aspects of temporal variability (heterochrony) after *Tc-zen2* RNAi that affect late morphogenesis. Our validation of specific transcriptional targets opens new avenues into serosal tissue biology and identifies a novel, paralogue-based regulatory circuit at the developmental transition from specification to maturation of the serosa. This now raises the question of how species with a single *zen* gene compare for the precision and progression of EEM development, and whether their molecular phenotypes support early Tc-Zen2 function as the consequence of iterative subfunctionalization events.

RESULTS

Recent tandem duplication of *zen* in the *Tribolium* lineage

We first surveyed *Tribolium* beetle genomes to assess sequence conservation at the *Hox3* locus. Using the *T. castaneum* paralogues as BLASTn queries, we find that the tandem duplication of *zen* is conserved across three closely related congenerics: *T. freemani*, *T. madens*, and *T. confusum* (Fig. 1A, 14-61 million years divergence [27]). Consistent with a recent event, phylogenetic analysis supports a single duplication at the base of the *Tribolium* lineage, and sequence alignments show particularly strong conservation in the homeobox, encoding the DNA-binding homeodomain (Figs. 1B, S1, S2).

Next, we investigated levels of coding sequence conservation between the *T. castaneum zen* (hereafter "*Tc-zen*") paralogues. Strongest nucleotide conservation occurs within the homeobox, where three conservation peaks correspond to the three encoded α-helices (Fig. 1C: >80% identity). In fact, within the coding sequence for the third α-helix there is a 20-bp stretch with 100% nucleotide identity (Fig. 1C), which is roughly the effective length of sequence for achieving systemic knockdown by RNA interference (RNAi) [28]. Indeed, *Tc-zen1*-specific double-stranded RNA (dsRNA) that spans the homeobox is sufficient to effect significant cross-paralogue knockdown of *Tc-zen2* (Fig. 1D: 19% of wild type expression with the long dsRNA fragment compared to only 35% with the short fragment; beta regression, $z=4.718$, $p<0.001$). As *Tc-zen2* is expressed in the tissue domain specified by Tc-Zen1 [26], it is in fact surprising that any residual *Tc-zen2* transcript is detected in the *Tc-zen1* RNAi background, an aspect we revisit below in our discussion on *zen* genes' regulation. Meanwhile, for our experimental design, we find that indeed a short fragment alone is sufficient to strongly knock down *Tc-zen1* itself (no significant change in knockdown efficiency between the long and short fragments: 10% and 11% expression, respectively; beta regression, $z=0.558$, $p=0.577$). For all subsequent paralogue-specific functional testing, we

thus designed our dsRNA fragments to exclude the homeobox and thereby avoid off-target effects (Fig. 1C: *Tc-zen1* short fragment: yellow; *Tc-zen2*: green).

Distinct roles of the *Tc-zen* paralogues at different developmental stages

EEM development has been well characterized morphologically in the beetle [13,29-31], including the *Tc-zen* paralogues' roles. Briefly, the first differentiation event distinguishes the serosa from the rest of the blastodermal cell sheet (Fig. 2A, at ~10% embryonic development). Tissue reorganization then involves serosal expansion and internalization of the embryo and amnion (EEM formation: subdivided into the "primitive pit" and "serosal window" stages). This topology is later reversed when the EEMs actively rupture and contract ("withdrawal"), coordinated with expansion of the embryo's flanks for dorsal closure of the body (Fig. 2C, at ~75% development). After *Tc-zen1* RNAi, presumptive serosal cells are respecified to other anterior fates, leading to an early enlargement of the head and amnion and loss of serosal identity (Fig. 2B) [26]. *Tc-zen2* RNAi impairs or wholly blocks late EEM withdrawal [12,26], confining the embryonic flanks such that the epidermis encloses the embryo's own legs instead of closing the back, leading to an everted (inside out) configuration (Fig. 2D) [12,32].

Here, we were able to reproduce the morphological phenotypes after RNAi for each *Tc-zen* paralogue (Fig. 2A′-D′). RNAi is particularly efficient for *Tc-zen1* (98.8% knockdown, Fig. 2E). Specific phenotypes after *Tc-zen2* RNAi (73.8% knockdown) include complete eversion (20.5%, Fig. 2D′) as well as milder defects in EEM withdrawal (53.3%, Figs. 2F, S3). Furthermore, we newly explored how the paralogues' functions relate to their transcript expression profiles across embryogenesis. Consistent with their functions, *Tc-zen1* has early expression while only *Tc-zen2* persists until the membrane rupture stage (Fig. 2G). Unexpectedly, late-acting *Tc-zen2* has peak transcript expression during early development.

The *Tc-zen* paralogues exhibit subtle differences in expression during early development

To gain insight into *Tc-zen* gene regulation and to determine the developmental stages of primary transcription factor function for each paralogue, we undertook a fine-scale spatiotemporal characterization of *Tc-zen1* and *Tc-zen2* expression for both transcript and protein (RT-qPCR, *in situ* hybridization, western blotting, immunohistochemistry).

As both paralogues are strongly expressed in early development (Fig. 2G), we first examined these stages in detail. *Tc-zen1* transcript arises in an anterior gradient during blastoderm formation (4-6 hours after egg lay, hAEL), peaks at the differentiated blastoderm stage with uniform expression throughout the presumptive serosa (6-10 hAEL), and then becomes patchy and retracts to a narrow region at the tissue's border during EEM formation (10-14 hAEL; Fig. 3A-F). After the EEMs have fully enclosed the early embryo, *Tc-zen1* transcript is no longer detected (Figs. 2G, 3A). Peak *Tc-zen1* transcript expression is followed shortly by detectable protein for Tc-Zen1, although this, too, only occurs during early development (Figs. 4A, S4, S5).

Tc-zen2 expression begins slightly later, at the differentiated blastoderm stage (6-8 hAEL), with peak levels occurring during EEM formation (10-14 hAEL; Fig. 3A). We also observed spatial differences between the paralogues. *Tc-zen2* is first detected only in an anterior subset of the serosa when *Tc-zen1* is expressed in the entire tissue (compare Fig. 3C,H). Then, *Tc-zen2* transcript expands throughout the serosa while *Tc-zen1* transcript retracts, concomitant with the expansion of the entire serosal tissue during EEM formation (compare Figs. 3D-F,I-K, S5). Notably, the *Tc-zen* paralogues are expressed consecutively, but not concurrently, at the rim of the serosa. It is only during late EEM formation that we first observe *Tc-zen2* expression throughout the entire serosal tissue (Fig. 3K). By this time, Tc-Zen2 protein is also strongly expressed and persists (Figs. 4A-B, S4, S5, and see below), while *Tc-zen2* transcript wanes gradually (from 14 hAEL; Fig. 3A).

Transcriptional impact of *Tc-zen1* and *Tc-zen2* during early embryogenesis

Since protein expression follows shortly after peak transcript expression for both paralogues (Figs. 3A, 4A), we used the high sensitivity of our RT-qPCR survey (Fig. 3A) to inform our staging for functional testing by RNAi. To identify transcriptional targets for each *zen* gene, our RNA-seq after RNAi approach assessed differential expression (DE) between age-matched wild type and knockdown samples. We focused specifically on the time windows of peak gene expression: 6-10 hAEL for *Tc-zen1* and 10-14 hAEL for *Tc-zen2* (curly brackets in Fig. 3A). These four-hour windows were chosen to maximize the number of identified target genes while prioritizing direct targets for Zen transcription factor binding.

The RNA-seq data are consistent with *a priori* expectations based on the morphological consequences of RNAi for each *zen* gene (Fig. 2A-D). That is, *Tc-zen1* has a clear early role in tissue specification, and its knockdown at these stages has a strong transcriptional impact, wherein principal component analysis (PCA) clearly distinguishes experimental treatments (Fig. 5A). In contrast, *Tc-zen2* has an early expression peak but its manifest role in late EEM withdrawal occurs nearly two days later (56% development later). Accordingly, we find a negligible effect on the early egg's total transcriptome after *Tc-zen2* RNAi (Fig. 5A), despite verification of efficient knockdown (Fig. 2F). RNAi efficiency was also confirmed directly with the RNA-seq data: both *Tc-zen1* and *Tc-zen2* exhibit DE reduction after their respective knockdown (Table S1A-B: significant fold change reductions of -4.86 for *Tc-zen1* and -3.54 for *Tc-zen2*). Overall, we obtained 338 DE genes after *Tc-zen1* RNAi compared to only 26 DE genes after *Tc-zen2* RNAi, while global transcriptional changes affect nearly 12% of all genes during early embryogenesis (2221 DE genes: Fig. 5F-a,c,d, Tables S1A-C).

Given the recent nature of the duplication, which is evident in the similarity of the *Tc-zen* paralogues' DNA-binding homeodomains and early expression profiles, we asked whether there is a legacy of shared early function. If this is the case, *Tc-zen2* might exhibit a subtle regulatory profile similar to *Tc-zen1*. However, even with relaxed thresholds for differential expression, we find few shared targets between the paralogues, particularly when the direction of regulation is considered (Fig. 5B, Tables S2A-B). Thus, we conclude that *Tc-*

zen2 has a minimal effect on early development, and that this does not constitute a transcriptional "echo" of co-regulation with *Tc-zen1* due to common ancestry. Why, then, is *Tc-zen2* strongly expressed during early development?

The *Tc-zen* paralogues are mutual regulatory targets

We next considered the *Tc-zen* paralogues as factors necessary for defining the serosal tissue, as indicated by their specific transcriptional targets. *Tc-zen1* is strictly required for serosal tissue identity [26]. Differentiation of the serosa involves an early switch from mitosis to the endocycle [29,30], resulting in polyploidy [13]. Consistent with this, we identified a homologue of the endocycle factor *fizzy-related* among DE genes upregulated by Tc-Zen1 (Table S1A) [33,34]. From known targets of Tc-Zen1, we also recovered *Dorsocross* and *hindsight*, involved in EEM formation [35], and *chitin synthase 1*, required for production of the protective cuticle [10]. Additionally, we hypothesized that the slight offset whereby *Tc-zen1* expression precedes *Tc-zen2* is consistent with *Tc-zen1* activating *Tc-zen2*. We could confirm this regulatory interaction both by RNA-seq and RT-qPCR after *Tc-zen1* RNAi (Fig. 6A-B). Thus, Tc-Zen1 as a serosal specifier upregulates factors for definitive tissue differentiation, including *Tc-zen2* as a candidate (Fig. 6I).

Are there Tc-Zen2 transcriptional targets that could support an early role in the serosa for this paralogue? Among the few genes with strong differential expression (Fig. 5F-d), we validated several as probable targets. These genes are expressed in the early serosa and/or their transcript levels are first strongly upregulated during peak *Tc-zen2* expression (12-14 hAEL; *e.g.*, Fig. S6). Their putative functions as enzymes or structural components for chitin-based cuticle (Cpr proteins) and as signaling molecules support a role for *Tc-zen2* in the physiological maturation of the serosa, providing complementary regulatory control to Tc-Zen1.

In performing reciprocal validation assays, we then uncovered an unexpected early function of Tc-Zen2 in the repression of its own paralogous activator. After *Tc-zen2* RNAi, *Tc-zen1* transcript is still detectable, a qualitative observation that previously had been taken as evidence for a lack of paralogue feedback [26]. The sensitivity of our quantification and spatial analyses across the dynamic early stages of development now allows us to correct this. We detect an upregulation of *Tc-zen1* that was only weakly suggested by our RNA-seq data but then strongly supported in RT-qPCR assays (Fig. 6A-B). We also confirmed this observation by *in situ* hybridization. After *Tc-zen2* RNAi, *Tc-zen1* transcript is expressed at higher levels than in wild type (compare Fig. 6C-D,F-G). *Tc-zen1* also remains strongly expressed throughout the serosa at stages when its wild type expression is restricted to low levels at the tissue rim (compare Fig. 6E,H). In fact, the abrupt reduction in *Tc-zen1* transcript and protein levels in wild type correlates with increasing *Tc-zen2* levels, and spatially their dynamic expression is largely complementary, if not outright mutually exclusive (Figs. 3,4). Together, these results suggest that Tc-Zen1 upregulates *Tc-zen2* in its wake, and that in turn early Tc-Zen2 represses *Tc-zen1*. Thus, the *Tribolium* paralogues

function as mutual regulatory targets, comprising an integrated regulatory module for early serosal development (Fig. 6I).

Tc-Zen2 is exclusively serosal, with persistent nuclear localization

To complete our analysis of *Tc-zen2*, we also examined its activity at later stages. We could detect both transcript (weakly, Figs. 2G, 3A) and protein (particularly strongly in mid-embryogenesis, Fig. 4A) continuously until the stage of EEM withdrawal, spanning 14-75% of development (10-54 hAEL, assayed in two-hour intervals; see also Fig. S4). Moreover, we find that Tc-Zen2 is persistently localized to the nucleus, demonstrated by fluorescent immunohistochemistry on cryosectioned material of selected stages (Fig. 4B-E,G,H). This contrasts with some species' orthologues, which show stage-specific exclusion of Zen protein from the nucleus [36]. We could also refine the spatial scope of *Tc-zen2* activity: in contrast to earlier reports [26], we found no evidence for *Tc-zen2* transcript or protein in the amnion (Fig. 4D-H″), indicating that this factor is strictly serosal.

Late transcriptional dynamics are largely serosa-specific and *Tc-zen2*-dependent

The early RNA-seq after RNAi experiment examined the time of peak *Tc-zen2* expression. Complementing this, we used the same approach to examine the stage of known *Tc-zen2* function in late EEM withdrawal. Withdrawal begins with rupture of the EEMs, at 52.1 ± 2.3 hAEL as determined by live imaging [31]. Here, we assayed the four-hour intervals just before (48-52 hAEL) and after (52-56 hAEL) rupture, to assess *Tc-zen2* transcriptional regulation that precedes and then accompanies withdrawal. Consistent with *Tc-zen2*'s known role, we detect >16× more DE genes after *Tc-zen2* RNAi in late development (>430 DE genes, compare Fig. 5F-e,f with 5F-d). PCA also clearly separates knockdown and wild type samples at late stages (Fig. 5C).

Our staging helps to contextualize *Tc-zen2* and EEM-specific processes relative to concurrent embryonic development. We thus evaluated differential expression in pairwise comparisons not only between wild type and RNAi samples, but also over time in both backgrounds (Fig. 5D,F-b,e-g). Comparisons across consecutive stages in early and late development (Fig. 5F-a,b) reveal two general changes in the wild type transcriptional landscape. There is far less dynamic change in gene expression in late development (5.8× fewer DE genes), consistent with steady state and ongoing processes in later embryogenesis compared to the rapid changes of early development. Also, whereas early development shows a fairly even balance between up- (48%) and downregulation (52%), late development is predominantly characterized by increasing expression levels over time (79%).

Against this backdrop, the transcriptional impact of *Tc-zen2* is quite pronounced. Most genes with changing expression over time in the late wild type background are also affected by *Tc-zen2* RNAi (Fig. 5D: 77%, 293/383 DE genes from green Venn diagram set). We detect this strong effect even though *Tc-zen2* is restricted to the serosa (Fig. 4), a tissue that ceased mitosis (Fig. 6I) and comprises only a small cell population within our whole-egg

samples. This suggests that most dynamic transcription in late development pertains to EEM morphogenesis, with the global transcriptional impact of *Tc-zen2* at these stages even greater than for *Tc-zen1* in early development (Fig. 5F-e,f, *cf.* 5F-c). Most candidate *Tc-zen2* targets are differentially expressed at a single stage (72%), although a substantial fraction (26%) exhibits consistent activation or repression, while an intriguing handful of genes shows changing, stage-specific regulation (Fig. 5E). These patterns imply that the persistent nuclear localization of Tc-Zen2 (Fig. 4) reflects active transcriptional control, not merely localization to the nucleus or DNA binding in a paused, non-functional state [37].

To characterize late Tc-Zen2 activity, we functionally annotated and validated candidate transcriptional targets. Gene ontology enrichment tests confirmed that ongoing cuticle regulation is a primary role, including remodeling as the serosa detaches from its own cuticle in preparation for withdrawal (Fig. S7, Table S4A-B). For validation, we selected a dozen genes based on known biological processes for tissue remodeling (*e.g.*, cytoskeleton and morphogenesis), prominent GO categories (*e.g.*, transmembrane transporters), and evidence of dynamic regulation (Figs. 5E, S8, Tables S5-S6). All tested candidates were confirmed by RT-qPCR (Fig. S8). This included two of the genes that are first activated and then repressed by Tc-Zen2, where both genes encode proteins with conserved domains of unknown function (Table S6). Notably, such dynamic targets include a selected subset of the *Osiris* multigene locus (Fig. S8, [38]). Lastly, we evaluated Tc-Zen2 regulation of serosal immune genes [7]. Although our samples were not pathogen challenged, we could detect expression for 83% of these genes (89 of 107 genes), with 20% showing differential expression after *Tc-zen2* RNAi (Table S3A-B). Thus, while Tc-Zen2 is not a global effector, it may regulate subsets of immune genes. Notably, transcripts of most serosal immune genes (87 genes) continue to be detected during withdrawal, supporting their expression as an inherent feature of the serosa – even when it is no longer a protective layer enclosing the embryo.

Evidence of variable developmental delay after *Tc-zen2* RNAi

The *Tc-zen2*^{RNAi} molecular phenotype also provides new insight into the physical phenotype of defective EEM withdrawal, suggesting that a variable, partial delay in preparatory transcriptional changes is the underlying cause.

Several observations are consistent with a delay. As noted above, all late RNA-seq biological replicates cluster by treatment in PCA. Interestingly, the older *Tc-zen2*^{RNAi} samples (52-56 hAEL) have intermediate component scores compared to the younger *Tc-zen2*^{RNAi} and younger wild type samples (48-52 hAEL, Fig. 5C). Similarly, DE comparisons identify noticeably fewer DE genes between the older *Tc-zen2*^{RNAi} sample and either of the younger samples (Fig. 5F-g,h, Tables S3D-E). In fact, the very low number of DE genes implies that there is virtually no difference in the transcriptional profile of the older *Tc-zen2*^{RNAi} sample compared to the younger wild type sample (Fig. 5F-h). Furthermore, nearly all genes that change in expression over time in the *Tc-zen2*^{RNAi} background are also candidate targets of Tc-Zen2 at the pre-rupture stage (95%, Fig. 5D: inset Venn diagram). In other words, *Tc-zen2*^{RNAi} eggs generally require an additional four hours (5.6% development) to attain a

transcriptional profile comparable to the wild type pre-rupture stage, and this is achieved by belated activation of Tc-Zen2 target genes. However, only a subset of genes exhibit delayed recovery (34%, Fig. 5D inset). These target genes may thus be independently activated by other factors, in addition to activation by Tc-Zen2.

Our RNA-seq data also reveal increased variability after *Tc-zen2* RNAi. The pre-rupture *Tc-zen2*^{RNAi} biological replicates show comparably tight clustering to their age-matched wild type counterparts (48-52 hAEL, Fig. 5C). This suggests that pre-rupture is the stage of primary Tc-Zen2 function, also supported by our detection of the greatest number of DE genes at this stage (compare Fig. 5F-d,e,f). In contrast, the older *Tc-zen2*^{RNAi} samples have a noticeably greater spread along the vectors of the first two principal components (52-56 hAEL, Fig. 5C), consistent with cumulative variability as the RNAi phenotype develops, presumably in part due to the observed partial transcriptional recovery (Fig. 5D). This variability may in itself provide explanatory power for the spectrum of end-stage *Tc-zen2*^{RNAi} phenotypes (Fig. 2F, see below).

DISCUSSION

Our analysis of regulation upstream and downstream of the beetle *zen* genes reveals several unexpected features regarding the evolution and biological roles of these unusual paralogues.

First, sequence conservation belies the extent of *zen* paralogue functional divergence. Fine-tuned transcriptional regulation is required to restrict regulatory crosstalk, and conserved non-coding regions may contribute to this. The region upstream of *zen1* has particularly high conservation and was tested as an *in vivo Tc-zen1* reporter in a recent study [39] (Fig. 1A: dashed line region). This construct recapitulates expression at the rim of the serosal window, a feature common to both paralogues (as in Figs. 3F,K, 6E). However, early blastoderm *Tc-zen1* expression is absent (*cf.*, Fig. 3B-C), while subsequent embryonic/amniotic expression of this construct represents a wholly ectopic domain [39]. Our own observations that detectable, residual levels of *Tc-zen2* persist after *Tc-zen1* RNAi (Fig. 1D: short fragment) and that low but detectable levels of *Tc-zen2* are expressed during blastoderm formation (4-6 hAEL, compare Fig. 3A and 3G) imply early *Tc-zen2* regulation that is independent of *Tc-zen1*, possibly involving unknown anterior terminal activators relevant for *Tc-zen1* itself. Thus, regulation of the *Tc-zen* genes requires multiple inputs that remain to be elucidated.

Specificity of regulation by the *Tc-zen* genes is also elusive. In the *Tc-zen* homeobox, sequence similarity is particularly high in the third α-helix, which confers DNA-binding specificity (Fig. 1C-D) [20,40,41]. Yet, the paralogues' shared ancestry is not reflected in redundant activity (Fig. 5A-B). Rather, strong conservation across *Tribolium* species, particularly of *zen2* (Fig. S1), may indicate not only limited divergence but also positive, purifying selection [42]. How, then, do the paralogues regulate different targets? In canonical Hox3 proteins, DNA-binding specificity can be enhanced by the common Hox co-factor Extradenticle [41]. In contrast, insect Zen proteins have lost the hexapeptide motif required for

this interaction, and no other co-factor binding motifs are known [16], deepening the long recognized "Hox specificity paradox" [43] in the case of the beetle *zen* paralogues. Recent work has revealed a role for differential chromatin accessibility in conferring Hox binding specificity [44]. This highlights an intriguing direction for future research, such as the extent of rapid functional genomic changes that may occur in early development. However, the stage-specific and gene-specific regulatory activity of Tc-Zen2 (Figs. 5E, S8) argues for greater regulatory precision than expected based on opportunistic binding to accessible targets alone. Meanwhile, our molecular dissection of the *Tc-zen* paralogues elucidates the strong extent of their functional divergence.

Specifically, mutual regulation has implications for the paralogues' network logic and confers temporal precision. The newly discovered negative feedback loop of Tc-Zen1 activation leading to repression by Tc-Zen2 constitutes a tight linkage. To what extent could *Tc-zen1* overexpression bypass upregulation of *Tc-zen2* as its target, resulting in repression of *Tc-zen1* and thus cancelling out the manipulation? In fact *Tc-zen2* RNAi does confer overexpression of *Tc-zen1* and reduced *Tc-zen2* (Fig. 6, Table S1B). However, consistently lower knockdown efficiency for *Tc-zen2* than for *Tc-zen1* (Fig. 2, [26]) may reflect a dose-limiting lack of regulatory disentanglement. Arguably, *Tc-zen1* and *Tc-zen2* together satisfy the criteria of a minimal gene regulatory network (GRN) kernel [45], including "recursive wiring" and the experimental challenges this entails. Alternatively, the *Tc-zen* paralogues could be viewed as a single unit in a serosal GRN and thus qualify as a "paradoxical component" that both activates and inhibits (Fig. 6I) [46]. Consistent with theoretical expectations, delayed inhibition produces a discrete pulse of *Tc-zen1* (Figs. 3,4). As the pulse is non-oscillatory, this may also imply that *Tc-zen2* is a positive autoregulator [46], a property known for a number of canonical Hox genes but thus far not known for insect *zen* genes (e.g., [47,48]).

Furthermore, Tc-Zen2 was previously implicated in the unusual role of translational repression of the early embryonic factor *caudal* [49]. Conceivably, Tc-Zen2 repression of *Tc-zen1* could act in a composite fashion, at both the transcriptional and translational levels. Composite activity could expedite repression – consistent with *Tc-zen1*'s abrupt decline (Figs. 3A, 4A) – and enhance stability of the system [50]. Similar regulatory dynamics have also been found in other contexts. Feedback loops with activation leading to inhibition can promote robustness [51] and rapidly, precisely restrict expression [52,53]. Whereas spatial precision contributes to patterning of distinct tissues [53], the serosal Zen feedback loop generates temporal precision.

In other words, paralogue divergence promotes the progression of serosal development. Negative feedback implies a strong developmental requirement to repress *Tc-zen1*, even before the serosa has fully enclosed the embryo. Since *Tc-zen2* persists in this same domain, why is this necessary? In *Drosophila*, expression of *Dm-zen* is also short-lived [54], and its overexpression causes an increase in amnioserosal cell and nuclear size [55]. The insect EEMs are known to be polyploid to characteristic levels [13,56,57], and excessive ploidy could interfere with the tissues' structure and function as barrier epithelia [58]. Our RNA-seq

DE analyses support a role for *Tc-zen1*, but not *Tc-zen2*, in promoting serosal endoreplication (Fig. 6I). Thus Tc-Zen2's repression of *Tc-zen1* may ensure a limited time window for this transition. The temporal offset also limits the amount of gene product of the paralogues' sole shared target, the cuticle maturation factor *aaNAT*, and overall effects temporally graded cuticle production (Figs. 5B,6I,S6). Thus, the distinct roles of the *Tc-zen* paralogues offer a novel opportunity for regulatory refinement in the early serosa, with a finely tuned genetic separation of specification and maturation functions that fosters developmental progression.

Stepping back to a macroevolutionary framework, what are the implications of the *Tc-zen* paralogues for the evolution of insect *zen*? Although a specification function has only been demonstrated in the Holometabola [17-19,26,55], early expression is also known from some hemimetabolous species [36,59]. This suggests that the ancestral *zen* may have fulfilled both early specification and late morphogenesis roles. Then, the prominent functions of the *Tc-zen* paralogues would represent an instance of subfunctionalization [60]. Furthermore, *Tc-zen1* and *Dm-zen* differ from all other known homologues in lacking persistent serosal expression. Although implications of *Dm-zen* temporal restriction have been extensively discussed [54,61], its downregulation is likely passive [62,63] and occurs much later than for *Tc-zen1*. Meanwhile, the *Tribolium* innovation of having two functional, early extraembryonic copies of *zen* may have originated as redundant early expression, but Tc-Zen2's repression of *Tc-zen1* constitutes a new regulatory link that ensures Tc-Zen2 is the sole functional *zen* gene for late morphogenesis. This repression thus subdivides a composite ancestral role between the paralogues: *Tc-zen1* for early specification, *Tc-zen2* for late morphogenesis. A second regulatory innovation, Tc-Zen1's activation of *Tc-zen2*, then conferred the temporal offset in early paralogue activity and resulted in a second, nested instance of subfunctionalization, for serosal specification (Fig. 6I-J). Thus, whereas loss of autoregulation may have contributed to subfunctionalization of certain Hox genes among the four paralogous Hox clusters of vertebrates [64], here we infer that the acquisition of paralogue mutual regulation drove active, iterative subfunctionalization.

We have uncovered multiple roles of *Tc-zen2* as a diverged Hox gene throughout the lifetime of the serosal tissue, itself a morphological innovation [5]. Early, Tc-Zen2's repression of *Tc-zen1* (Fig. 6) and *Tc-caudal* [49] is noteworthy. A predominantly repressive role contrasts with Hox genes typically serving as activators, as do both *Tc-zen* paralogues at the stages of their primary function (Fig. 5F-c,e). Also, the precise mechanism and targets of potential Tc-Zen2 translational repression remain open questions. Future work will clarify whether such a function arose independently in *Tribolium* Zen2 and dipteran Bicoid [20,23,40] as distinct *Hox3/zen* derivatives. In later development, the serosa is the cellular interface with the outer environment. Our data elucidate Tc-Zen2's roles in the known protective functions of cuticle formation [10] and innate immunity [7]. Beyond this, our DE gene sets comprise a large, unbiased sample of candidate targets, laying the foundation for investigating wider roles of *Tc-zen2* in this critical tissue.

Finally, we identify *Tc-zen2*-dependent EEM withdrawal as the major transcriptionally regulated event in late embryogenesis and assess its precision (Fig. 5D).

Temporal and molecular variability after *Tc-zen2* RNAi underpins observed variability in knockdown EEM tissue structure, integrity, and morphogenetic competence, defining the broad spectrum of end-stage phenotypes (Figs. 2, S3). This ranges from mild defects in dorsal closure after transient EEM obstruction to persistently closed EEMs that cause complete eversion of the embryo (Figs. 2, S3, [12]). The unifying feature is a heterochronic shift of extraembryonic compared to embryonic developmental processes (delayed EEM withdrawal compared to epidermal outgrowth for dorsal closure).

There may also be species-specific differences in the timing of Zen function for withdrawal morphogenesis. The sole *zen* orthologue in the milkweed bug *Oncopeltus fasciatus* has a similarly persistent expression profile and specific role in withdrawal, termed "katatrepsis" in this and other hemimetabolous insects [17]. We previously observed a number of *Of-zen*-dependent, long-term morphological changes prior to rupture [65], contrasting with the more proximate effect of *Tc-zen2* (discussed above). Taking the work forward, it will be interesting to compare *Tc-zen2* and *Of-zen* transcriptional targets. Evaluating conserved regulatory features of EEM withdrawal across the breadth of the insects will clarify macroevolutionary patterns of change in the very process of epithelial morphogenesis.

METHODS

Tribolium castaneum stock husbandry

All experiments were conducted with the San Bernardino wild type strain, maintained under standard culturing conditions at 30 °C and 40-60% relative humidity [66].

In silico analyses

Draft genome assemblies for *T. freemani*, *T. madens*, and *T. confusum* were obtained as assembled scaffolds in FASTA-format (version 26 March 2013 for each species), accessed from the BeetleBase.org FTP site at Kansas State University (ftp://ftp.bioinformatics.ksu.edu/pub/BeetleBase/). Transcripts for *Tc-zen1* (TC000921-RA) and *Tc-zen2* (TC000922-RA) were obtained from the *T. castaneum* official gene set 3 (OGS3)[67]. These sequences were used as queries for BLASTn searches in the other species' genomes (BLAST+ 2.2.30) [68,69]. Sequences were extracted to comprise the *Hox3/zen* genomic loci, spanning the interval from 5 kb upstream of the BLASTn hit for the 5′ UTR of *Tc-zen1* to 5 kb downstream of the BLASTn hit for the 3′ UTR of *Tc-zen2*. These genomic loci were then aligned with the mVista tool [70,71] using default parameters. Nucleotide identities were calculated for a sliding window of 100 bp.

The maximum likelihood phylogenetic tree (Fig. 1B) was constructed based on an alignment of full-length Zen proteins, with gaps permitted, using the Phylogeny.fr default pipeline settings, with MUSCLE v3.8.31 alignment and PhyML v3.1 phylogenetic reconstruction [72]. The same topology and comparable support values were also obtained with additional sequences and other methods (Fig. S2). This includes *Drosophila* Zen and/or Z2 and/or *Drosophila* and *Megaselia* Bicoid, and/or insect Zen proteins with known expression but uncharacterized function [36,59], and bioinformatic predictions of Zen proteins in newly

sequenced genomes. This also holds for trees generated with Bayesian methods from the same interface (MrBayes program v3.2.6, 1000 generations, 100 burn-in trees). That is, the *Tribolium* Zen proteins form a clear clade with various Zen outgroups, and then with the fly proteins as long branch outgroups.

Coding sequence for the *Tc-zen* paralogues was aligned with ClustalW [73], with manually curation to ensure a gap-free alignment of the homeobox. Nucleotide identities were calculated for a sliding window of 20 bp, using Simple Plot [74].

RT-qPCR

RNA was extracted using TRIzol Reagent (Ambion) according to the manufacturer's protocol. RNA quality was assessed by spectrophotometry (NanoDrop 2000, Thermo Fisher Scientific). cDNA was synthesized using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). RT-qPCR was performed as described [35], using SYBR Green Master Mix (Life Technologies) and GoTaq qPCR Master Mix (Promega), with *Tc-RpS3* as the reference gene. Note that for *Tc-zen2* more consistent results were obtained using SYBR Green Master Mix. "Relative abundance" was calculated for each sample as the ratio relative to a pooled template control with cDNA from all depicted samples (method as in [35]). Samples were measured for the *Tc-zen* paralogues' wild type expression profiles (four biological replicates: Figs. 2G,3A) and evaluation of knockdown strength (three biological replicates: Figs. 1D,6B,S8). Intron-spanning primers were used for each *Tc-zen* paralogue and the selected candidate target genes (Table S7).

Parental RNAi and knockdown assessments

Parental RNAi was performed as described [26], with dsRNA synthesized with specific primers (Table S7) and resuspended in double-distilled water (ddH_2O). Generally, 0.3-0.4 µg of dsRNA was used to inject one pupa.

Analysis of knockdown efficiency with different *Tc-zen1* dsRNA fragments involved statistical tests on RT-qPCR data. The strength of the *Tc-zen* paralogues' knockdown using short and long *Tc-zen1* dsRNA fragments (Fig. 1C-D) was tested with a beta regression analysis in R v3.3.2 [75] using the package betareg v3.1-0 [76]. Relative expression of the *Tc-zen* paralogues in knockdown samples relative to wild type was used as the response variable and dsRNA fragment length as the explanatory variable.

For $Tc\text{-}zen1^{RNAi}$ phenotypic scoring (Fig. 2E), serosal cuticle presence/absence was determined by piercing the fixed, dechorionated egg with a disposable needle (Braun Sterican 23G, 0.60 x 25 mm): mechanically resistant eggs were scored for presence of the serosal cuticle while soft eggs that collapsed lacked serosal cuticle.

For $Tc\text{-}zen2^{RNAi}$ phenotypic scoring, larval cuticle preparations (Figs. 2C′,D′,F, S3) were produced as previously described [26].

Histology: *in situ* hybridization, cryosectioning, immunohistochemistry

Whole mount *in situ* hybridization was performed as described [31], with probes synthesized from gene specific primers (Table S7) and colorimetric detection with NBT/BCIP. Specimens were imaged in Vectashield mountant with DAPI (Vector Laboratories) for

nuclear counterstaining. Images were acquired on an Axio Plan 2 microscope (Zeiss). Image projections were generated with AxioVision (Zeiss) and HeliconFocus 6.7.1 (Helicon Soft).

For cryosectioning, embryos were embedded in liquid sucrose-agarose embedding medium (15% sucrose, 2% agarose, [my-Budget Universal Agarose, Bio-Budget], PBS). Solid blocks of embedding medium containing embryos were stored overnight in 30% sucrose solution in PBS at 4 °C. The blocks were then embedded in Tissue Freezing Medium (Leica Biosystems) and flash-frozen in ice-cold isopentene (2-methylbutane). Samples were serially sectioned (20 µm, longitudinal; 30 µm, transverse) with a CM1850 cryostat (Leica Biosystems).

Protein was detected for both Tc-Zen1 and Tc-Zen2 with specific peptide antibodies (gift from the laboratory of Michael Schoppmeier) [77]. Immunohistochemistry on whole mounts and on sectioned material was performed by washing the samples six times for 10 min. in blocking solution (2% BSA, 1% NGS, 0.1% Tween-20, PBS) followed by overnight incubation with the first antibody (rabbit anti-Tc-Zen1 and anti-Tc-Zen2, 1:1,000) at 4 °C. Next, the samples were washed six times for 10 min. in the blocking solution, followed by incubation with the secondary antibody (anti-rabbit Alexa Fluor 488 conjugate, 1:400, Invitrogen) for 3 h at room temperature (RT). Last, the samples were washed six times for 10 min. in the blocking solution. Samples were then mounted in Vectashield mountant with DAPI. Low magnification images were acquired with an Axio Imager 2 equipped with an ApoTome 2 (Zeiss) structured illumination module, and maximum intensity projections were generated with ZEN blue software (Zeiss). High magnification images were acquired with an LSM 700 confocal microscope (Zeiss) and the projections were generated with ZEN 2 black software (Zeiss).

Western blots

For each two-hour developmental interval, protein lysate was obtained by lysing whole eggs in RIPA buffer (150 mM NaCl, 1.0% IGEPAL® CA-630, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris, pH 8.0; Sigma-Aldrich catalogue # R0278), followed by centrifugation. For each sample, 50 µg of protein extract was separated by SDS-PAGE. The PageRuler Prestained Protein Ladder (10 to 180 kDa; Thermo Fisher Scientific, catalogue # 26616) was used as a size standard. Separated proteins were transferred onto nitrocellulose membrane (Thermo Fisher Scientific), which was blocked for 1 h in the blocking solution (100 mM Tris, 150 mM NaCl, pH 7.5, 0.1% Tween-20, 3% milk powder [Bebivita, Anfangsmilch]). Next, the membrane was incubated overnight at 4 °C with the first antibody (rabbit anti-Tc-Zen1 and anti-Tc-Zen2 as for immunohistochemistry, 1:1,000; mouse anti-Tubulin [Sigma-Aldrich #T7451: Monoclonal anti-acetylated tubulin], 1:10,000). Afterwards, the membrane was washed three times for 10 min. with the blocking solution at RT. The membrane was then incubated with the secondary antibodies (anti-rabbit and anti-mouse, HRP, 1:10,000, Novex) for 1 h at RT. Last, after the membrane was washed three times for 10 min. with the blocking solution at RT, the membrane was incubated with ECL substrate according to the manufacturer´s protocol (WesternSure ECL Substrate, LI-COR) and digital detection was performed on a western blot developing machine (C-DIGIT, LI-COR) with the high sensitivity settings. While the predicted molecular weights are 28.4 kDa for Tc-Zen1 and 33.7 kDa for Tc-Zen2 (ExPASy Mw Tool [78], last accessed 30 March 2020), in our hands we

consistently observed migration of our samples with actual sizes of approximately 39 kDa for Tc-Zen1 and 42 kDa for Tc-Zen2, which may reflect post-translational modifications.

RNA-sequencing after RNAi

RNAi for RNA-seq used the short, paralogue-specific dsRNAs (depicted schematically in Fig. 1C). For transcriptomic profiling by RNA-seq, three separate *Tc-zen1*^{RNAi} experiments (each with a separate population of injected females) were conducted. A total of seven *Tc-zen2*^{RNAi} experiments were conducted: one for each biological replicate at each developmental stage. Samples chosen for sequencing were assessed by RT-qPCR for level of knockdown in RNAi samples, with *Tc-zen1* reduced to ~10% of wild type levels and *Tc-zen2* to ~24% across biological replicates. For early development (6-14 hAEL), three biological replicates were sequenced for each experimental treatment, with 100-bp paired end reads on an Illumina HiSeq2000 machine. For late development (48-56 hAEL), four biological replicates were sequenced with 75-bp paired end reads on a HiSeq4000 machine. All sequencing was performed at the Cologne Center for Genomics (CCG), with six (HiSeq2000) or eight (HiSeq4000) multiplexed samples per lane yielding ≥6.6 Gbp per sample. The complete dataset from all 56 RNA-seq samples will be deposited in GenBank under accession number XXXX.

The quality of raw Illumina reads was examined with FastQC [79], and all RNA-seq samples were retained for analysis. The adaptor sequences and low quality bases were removed with Trimmomatic v0.36 [80]. Trimmomatic was also used to shorten 100-bp reads from the 3´ end to 75-bp reads to increase mapping efficiency (Table S8) [81]. The overrepresented sequences of mitochondrial and ribosomal RNA were filtered out by mapping to a database of 1266 *T. castaneum* mitochondrial and ribosomal sequences extracted from the NCBI nucleotide database (accessed 21 October 2016, search query "'tribolium [organism] AND (ribosomal OR mitochondrial OR mitochondrion) NOT (whole genome shotgun) NOT (Karroochloa purpurea)") with Bowtie2 v2.2.9 [82]. Trimmed and filtered reads were mapped to the *T. castaneum* OGS3 [67] (file name: Tcas5.2_GenBank.corrected_v5.renamed.mrna.fa) with RSEM [83]. The raw read count output from RSEM was compiled into count tables.

Both principal component and differential expression analyses were performed in R using the package DESeq2 v1.14.1 [84] with default parameters. For PCA, raw (unfiltered) read counts were used. For DE analyses, to eliminate noise all genes with very low read counts were filtered out by sorting in Microsoft Excel, following recommendations [85]. Specifically, genes were excluded from DE analysis if read counts ≤10 in ≥1 biological replicates for both the knockdown and wild type samples. For serosal immune genes, a given gene was considered to be expressed in the late serosa if we detected ≥100 RNA-seq reads in each of the four biological replicates for our wild type samples. Note that, throughout, our reporting of "DE genes" refers to analyses across all isoforms (18,536 isoform models) in the *T. castaneum* official gene set OGS3 [67].

Gene ontology (GO) analyses

GO enrichment analysis was performed by Blast2GO [86] using two-tailed Fisher´s exact test with a threshold false discovery rate (FDR) of 0.05.

GO term analysis was performed by Blast2GO against the *Drosophila* database (accessed 9 June 2017). Only GO terms from the level 5 were considered. Next, GO terms were grouped into categories of interest based on similarity in function (Table S5). Afterwards a unique count of *T. castaneum* gene sequences was calculated for each category of interest and the percentage was compared to the rest of the GO terms in the level 5 for each GO domain (Fig. S8).

DECLARATIONS

ACKNOWLEDGMENTS
We thank Denise Mackrodt and Michael Schoppmeier for the kind gift of the Tc-Zen1 and Tc-Zen2 peptide antibodies, Viera Kovacova for bioinformatic program recommendations, Luigi Pontieri for assistance with statistical analyses, Thorsten Horn for sharing unpublished data on cuticle gene expression, Gustavo Lazzaro Rezende for discussions on cuticle regulation, and Hilary Ashe and Chris Rushlow for insights into *zen* regulation in *Drosophila*. We also thank Miltos Tsiantis and Siegfried Roth for helpful discussions and recommendations throughout the course of this research project. Siegfried Roth, Peter Heger, and Matthias Pechmann provided helpful feedback on the manuscript.

FUNDING
This work was supported by funding from the German Research Foundation (Deutsche Forschungsgemeinschaft) through SFB 680 project A12 and Emmy Noether Program grant PA 2044/1-1 to KAP.

AUTHOR CONTRIBUTIONS
DG designed experiments, collected and analyzed data, established the bioinformatic pipeline for the RNA-seq data, wrote the paper.
IMVJ analyzed data, established the bioinformatic pipeline for the RNA-seq data, edited the manuscript.
KAP conceived the project, designed experiments, analyzed data, established the bioinformatic pipeline for the RNA-seq data, wrote the paper.

AVAILABILITY OF DATA AND MATERIALS
All RNA-seq read data are publically deposited at GenBank (NCBI), under accession number XXXX. The paralogue-specific peptide antibodies are available on request from the authors or from the source laboratory [77].

COMPETING INTERESTS
The authors declare that they have no competing interests

Figure 1. High sequence conservation of *Tribolium zen* orthologues and paralogues.
(A) Sequence conservation at the *Hox3/zen* locus of four *Tribolium* congenerics (represented by four-letter species abbreviations), using the *T. castaneum* locus as the reference sequence, assayed with a 100-bp sliding window. Non-coding sequence (pink) includes a highly conserved region that was recently tested as a *Tc-zen1* reporter (dashed line, [39], see Discussion).
(B) (Left) Representative phylogeny supporting a lineage-specific clade for the *Tribolium* Zen proteins (see Methods, Fig. S2), with single-copy orthologue outgroups for the specification and morphogenesis functions (*Megaselia abdita* [18] and *Oncopeltus fasciatus* [17], respectively). Node support values (%) are indicated; branch length unit is substitutions per site. (Right) Pie chart summaries of substitutions at the nucleotide (Nt.) and amino acid (AA) levels within the homeobox of *Tribolium zen* orthologues (see also Fig. S1).
(C) *T. castaneum zen* paralogue coding sequence comparison, based on a 20-bp sliding window. The final alignment of 887 positions includes 399 identities and 154 unaligned positions (gaps). Colored lines denote the homeobox and dsRNA fragments (see legend).
(D) *Tc-zen1* RNAi with the long dsRNA fragment causes cross-paralogue knockdown, with significantly stronger reduction in *Tc-zen2* levels (beta regression tests: see main text). Mean expression levels (RT-qPCR) are shown from three biological replicates after RNAi with the indicated dsRNA fragments (mapped in panel C, same color code); error bars represent the upper and lower quantiles (lower=0.025, upper=0.975).

Figure 2. *Tc-zen* paralogue roles in early specification (*Tc-zen1*) or late EEM morphogenesis (*Tc-zen2*).
(A-D´) Phenotypic consequences of RNAi for the *Tc-zen* paralogues are depicted: schematically at the time of primary defect (A-D, upper row) and with micrographs for the resulting phenotypes (A′-D′, lower row: DAPI nuclear stain (A′-B′) and cuticle preparations (C′-D′); see also Fig. S3). The dashed lines in the schematics in A,B denote the anterior embryonic border. Anatomical abbreviations: a, antenna; h, head; l, leg; t, telson; t3, third thoracic segment). Scale bars are 100 µm.
(E-F) RNAi phenotypic penetrance per paralogue. Sample sizes are numbers of embryos.
(G) Expression profiles of *Tc-zen1* and *Tc-zen2* in early (6-14 hAEL) and late (42-52 hAEL) development (RT-qPCR). Mean expression levels are shown from four biological replicates; error bars represent one standard deviation. Staging abbreviations: BF, blastoderm formation / uniform blastoderm; DB, differentiated blastoderm; PP, primitive pit; SW, serosal window; GBE, extended germband; pre-R, pre-rupture; MR, extraembryonic membrane rupture. Time is hours after egg lay (hAEL).

Figure 3. Transcript expression dynamics of the *Tc-zen* paralogues during early embryogenesis.
(A) Quantification of transcript levels in two-hour intervals (RT-qPCR, as in Fig. 2G).
(B-K´) Whole mount *in situ hybridization* for *Tc-zen1* (B-F) and *Tc-zen2* (G-K), with nuclear counterstains for morphological staging (B´-K´, arrows label the expanding serosal border).

All micrographs are oriented with anterior left and shown in lateral aspect with dorsal up (except in B and G, which depict stages before this can be determined). Scale bars in B and G are 100 µm and apply to B-F′ and G-K′, respectively.

Figure 4. Paralogue-specific Tc-Zen protein expression time courses and spatial restriction of Tc-Zen2.
(A) Tc-Zen1 and Tc-Zen2 expression profiles, spanning 4-54 hAEL (western blots, with anti-Tubulin as an internal loading control at 59.0 kDa; left-hand size standards are in kDa). Each blot shows a chronological assay in 2-hour intervals (labels indicate the minimum age from each interval). Tc-Zen1 protein (upper blots, arrow) is briefly detectable, from blastoderm differentiation until early serosal window closure. Tc-Zen2 (lower blots, arrows) arises at the primitive pit stage and persists until membrane rupture. See also Fig. S4 for additional blots. Staging abbreviations as in Figs. 2-3, and: GBR, germband retraction.
(B-H) Staining in whole mount (B,C,F) and cryosectioned (D,E,G,H) preparations shows that *Tc-zen2* is specific to the serosa, with protein consistently localized to the nucleus. We found no evidence for *Tc-zen2* expression in the second extraembryonic tissue, the amnion. Tc-Zen2-negative tissue and nuclei of the amnion are labeled with arrows (F,F′,H′,H′′) or a dashed outline (G′). The specimen shown in panel F, where the serosa was removed prior to staining, derives from a staining experiment spanning all early developmental stages, and where the serosa was specifically stained for *Tc-zen2* transcript in other specimens, as seen in Fig. 3H-K. Images labeled with the same letter are of a single embryo. Nuclear counterstains are shown in magenta. Scale bars are as indicated: 100 µm (whole mounts and overviews of longitudinal sections), 50 µm (transverse sections and G′), 10 µm (insets and H′,H′′ focusing on nuclear localization).

Figure 5. *Tc-zen* genes' transcriptional control: relative impact, overlap, and stage-specific dynamics revealed by RNA-seq after RNAi.
(A) PCA score plot for *Tc-zen1* and *Tc-zen2* and their respective age-matched wild type samples in early development, at the stages of peak expression (three biological replicates).
(B) Venn diagram assessments of shared early targets of the *Tc-zen* paralogues, with high ($P_{adj}\leq 0.01$, $|FC|\geq 2$) or low ($P_{adj}\leq 0.05$, $|FC|>1$) stringency DE criteria. Subset pie chart summarizes the direction of regulation for shared targets (see legend). *Tc-zen2* appears as a shared high-stringency target, reflecting its direct experimental knockdown by *Tc-zen2* RNAi as well as Tc-Zen1's role in its endogenous upregulation (see below).
(C) PCA score plot of *Tc-zen2*^{RNAi} and wild type samples in late development, at the stages before and during EEM withdrawal (four biological replicates).
(D) Venn diagrams of numbers of late DE genes across pairwise comparisons (see legend, with total DE gene counts per comparison listed parenthetically; see also panel 5F-b,e,f,g).
(E) Detailed comparisons of DE genes by stage (pre-rupture, withdrawal) and direction of regulation (activation, inhibition). Values are the number of DE genes. See also panel 5F-e,f.
(F) Summary metrics indicate the number of DE genes per pairwise comparison, and the proportion that are activated or inhibited by a given *Tc-zen* paralogue or that increase or

decrease in expression over time in a given background (WT or RNAi). DE cut-off values are $P_{adj} \leq 0.01$, $|FC| \geq 2$. For complete gene lists and statistics, see Tables S1A-C and S3A-E. Treatment abbreviations: WT, wild type; *z1*, *Tc-zen1*^{RNAi}; *z2*, *Tc-zen2*^{RNAi}. Staging for all panels is summarized in the lower right legend.

Figure 6. Evidence and implications of a *zen* paralogue regulatory module.
(A-B) *Tc-zen1* and *Tc-zen2* as candidate transcriptional targets of the respective paralogue, determined in comparisons of wild type and RNAi samples at the stages of peak paralogue expression (see Fig. 3A). Regulatory changes are corroborated by three biological replicates each in assays from RNA-seq (A: mean fold change, P_{adj}) and RT-qPCR (B: mean relative expression, error bars represent one standard deviation).
(C-H´) Whole mount *in situ* hybridization for *Tc-zen1* in wild type (C-D) and after *Tc-zen2* RNAi (F-H), with nuclear counterstains for morphological staging (C´-H´). Wild type and RNAi eggs were stained for the same duration. All micrographs are oriented with anterior left and shown in lateral aspect with dorsal up, unless stated otherwise. Scale bar in C is 100 µm and applies to all micrographs.
(I) The *Tc-zen* paralogues comprise an integrated "*zen*" regulatory module for early serosal development. Paradoxical regulation (red) involves competing inputs to the same target. Composite regulation (blue) denotes potential repression at both the transcriptional and translational levels. Gene abbreviations correspond to the full names and descriptions in the main text.
(J) Evolutionary scenario for the ancestral extraembryonic functions of Zen and its subsequent changes in *Tribolium*, progressively partitioning functions via iterative subfunctionalization, enabled by the new paralogue cross-talk of mutual regulation. Primary subfunctionalization separated early specification (sp) and late morphogenesis (mo) functions. Secondly, the original specification function has now been finely subdivided into the initial, substantial role that Tc-Zen1 continues to play ("sp1") and the slightly later, subtle functions of Tc-Zen2 that we reveal here ("sp2").

SUPPLEMENTARY FILES

PDF:

Figure S1. Alignments of zen1 and zen2 homeoboxes/domains. (Supports Fig. 1B,C)
Figure S2. Additional phylogenies of insect class 3 Hox proteins. (Supports Fig. 1B)
Figure S3. Phenotypic consequences of Tc-zen2RNAi - milder defects. (Supports Fig. 2D,F)
Figure S4. Tc-Zen1 and Tc-Zen2 expression detected by western blotting. (Supports Fig. 4A)
Figure S5. Early *in vivo* expression of transcript and protein for the *Tc-zen* paralogues. (Supports Fig. 4)
Figure S6. Evaluation of Tc-Zen2 candidate targets. (Supports Fig. 5B)
Figure S7. Gene ontology (GO) enrichment analysis after Tc-zen2RNAi. (Supports Fig. 5D-E)
Figure S8. Functional annotation and validation of candidate late Tc-Zen2 targets. (Supports Fig. 5D-E)

Table S5. List of gene ontology (GO) terms and categories of interest. (Supports Figs. 5D, S8)
Table S6. Tc-Zen2 late candidate targets evaluated by RT-qPCR. (Supports Figs. 5D, S8)
Table S7. *Tribolium castaneum* (*TC*) primer sequences. (Supports Methods/ Figs. 1-2)
Table S8. Comparison of alignment statistics by read length. (Supports Methods/ Fig. 5)

Excel workbook with large datasets and gene lists:

Table	Description		Corresponding (main) text element
S1A	List of differentially expressed genes during early development after *Tc-zen1* RNAi (P_{adj}≤0.01)		Fig. 5F-c
S1B	List of differentially expressed genes during early development after *Tc-zen2* RNAi (P_{adj}≤0.01)		Fig. 5F-d
S1C	List of differentially expressed genes during early development (WT *Tc-zen1* peak, 6-10 hAEL vs. WT *Tc-zen2* peak, 10-14 hAEL) (P_{adj}≤0.01)		Fig. 5F-a
S2A	Regulatory targets shared by both *Tc-zen* paralogues with relaxed thresholds (P_{adj}≤0.05): same direction of regulation, N=42		Fig. 5B
S2B	Regulatory targets shared by both *Tc-zen* paralogues with relaxed thresholds (P_{adj}≤0.05): opposite direction of regulation, N=78		Fig. 5B
S3A	List of differentially expressed genes during late development (pre-rupture, 48-52 hAEL) after *Tc-zen2* RNAi (P_{adj}≤0.01)		Fig. 5F-e
S3B	List of differentially expressed genes during late development (rupture/post-rupture, 52-56 hAEL) after *Tc-zen2* RNAi (P_{adj}≤0.01)		Fig. 5F-f
S3C	List of differentially expressed genes during late development (WT pre-rupture, 48-52 hAEL vs. WT rupture/post-rupture, 52-56 hAEL) (P_{adj}≤0.01)		Fig. 5F-b
S3D	List of differentially expressed genes during late development (*Tc-zen2* RNAi pre-rupture, 48-52 hAEL vs. *Tc-zen2* RNAi rupture/post-rupture, 52-56 hAEL) (P_{adj}≤0.01)		Fig. 5F-g
S3E	List of differentially expressed genes during late development (WT pre-rupture, 48-52 hAEL vs. *Tc-zen2* RNAi rupture/post-rupture, 52-56 hAEL) (P_{adj}≤0.01)		Fig. 5F-h
S4A	Enriched GO terms in the dataset "Differentially expressed genes during pre-rupture (48-52 hAEL stage) after *Tc-zen2* RNAi"		Figs. 5D, S6A
S4B	Enriched GO terms in the dataset "Differentially expressed genes during rupture/post-rupture (52-56 hAEL stage) after *Tc-zen2* RNAi"		Figs. 5D, S6A

REFERENCES

1 Krumlauf, R. Evolution of the vertebrate Hox homeobox genes. *BioEssays* **14**, 245-252 (1992).

2 Armisen, D. *et al.* The genome of the water strider *Gerris buenoi* reveals expansions of gene repertoires associated with adaptations to life on the water. *BMC Genomics* **19**, 832 (2018).

3 McKenna, D. D. *et al.* Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. *Genome Biol.* **17**, 227 (2016).

4 Negre, B. & Ruiz, A. HOM-C evolution in *Drosophila*: is there a need for *Hox* gene clustering. *Trends Genet.* **23**, 55-59 (2007).

5 Panfilio, K. A. Extraembryonic development in insects and the acrobatics of blastokinesis. *Dev. Biol.* **313**, 471-491 (2008).

6 Chen, G., Handel, K. & Roth, S. The maternal NF-kappaB/Dorsal gradient of *Tribolium castaneum*: dynamics of early dorsoventral patterning in a short-germ beetle. *Development* **127**, 5145-5156 (2000).

7 Jacobs, C. G. C., Spaink, H. P. & van der Zee, M. The extraembryonic serosa is a frontier epithelium providing the insect egg with a full-range innate immune response. *eLife* **3**, e04111 (2014).

8 Cobben, R. H. *Evolutionary Trends in the Heteroptera, Part I. Eggs, Architecture of the Shell, Gross Embryology and Eclosion*. (Centre for Agricultural Publishing and Documentation, 1968).

9 Rezende, G. L. *et al.* Embryonic desiccation resistance in *Aedes aegypti*: presumptive role of the chitinized serosal cuticle. *BMC Dev. Biol.* **8**, 82 (2008).

10 Jacobs, C. G. C., Rezende, G. L., Lamers, G. E. M. & van der Zee, M. The extraembryonic serosa protects the insect egg against desiccation. *Proc. R. Soc. B* **280**, 20131082 (2013).

11 Farnesi, L. C., Menna-Barreto, R. F. S., Martins, A. J., Valle, D. & Rezende, G. L. Physical features and chitin content of eggs from the mosquito vectors *Aedes aegypti, Anopheles aquasalis* and *Culex quinquefasciatus*: Connection with distinct levels of resistance to desiccation. *J. Insect Physiol.* **83**, 43-52 (2015).

12 Hilbrant, M., Horn, T., Koelzer, S. & Panfilio, K. A. The beetle amnion and serosa functionally interact as apposed epithelia. *eLife* **5**, e13834 (2016).

13 Panfilio, K. A., Oberhofer, G. & Roth, S. High plasticity in epithelial morphogenesis during insect dorsal closure. *Biol. Open* **2**, 1108-1118 (2013).

14 Horn, T., Hilbrant, M. & Panfilio, K. A. Evolution of epithelial morphogenesis: phenotypic integration across multiple levels of biological organization. *Front. Genet.* **6**, 303 (2015).

15 Ferguson, L. *et al.* Ancient expansion of the Hox cluster in Lepidoptera generated four homeobox genes implicated in extra-embryonic tissue formation. *PLoS Genet.* **10**, e1004698 (2014).

16 Panfilio, K. A. & Akam, M. A comparison of Hox3 and Zen protein coding sequences in taxa that span the *Hox3/zen* divergence. *Dev. Genes Evol.* **217**, 323-329 (2007).

17 Panfilio, K. A., Liu, P. Z., Akam, M. & Kaufman, T. C. *Oncopeltus fasciatus zen* is essential for serosal tissue function in katatrepsis. *Dev. Biol.* **292**, 226-243 (2006).

18 Rafiqi, A. M., Lemke, S., Ferguson, S., Stauber, M. & Schmidt-Ott, U. Evolutionary origin of the amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of *zen*. *Proc. Natl Acad. Sci. USA* **105**, 234-239 (2008).

19 Wakimoto, B. T., Turner, F. R. & Kaufman, T. C. Defects in embryogenesis in mutants associated with the Antennapedia Complex of *Drosophila melanogaster*. *Dev. Biol.* **102**, 147-172 (1984).

20 McGregor, A. P. How to get ahead: the origin, evolution and function of *bicoid*. *BioEssays* **27**, 904-913 (2005).

21 Pultz, M. A., Diederich, R. J., Cribbs, D. L. & Kaufman, T. C. The proboscipedia locus of the Antennapedia complex: a molecular genetic analysis. *Genes Dev.* **2**, 901-920 (1988).

22 Rushlow, C. & Levine, M. Role of the *zerknüllt* gene in dorsal-ventral pattern formation in *Drosophila*. in *Advances in Genetics: Genetic Regulatory Hierarchies in Development* Vol. 27 *Advances in Genetics* (ed Theodore R. F. Wright) (Academic Press, 1990).

23 Stauber, M., Jäckle, H. & Schmidt-Ott, U. The anterior determinant *bicoid* of *Drosophila* is a derived class 3 Hox gene. *Proc. Natl Acad. Sci. USA* **96**, 3786-3789 (1999).

24 Falciani, F. *et al.* Class 3 Hox genes in insects and the origin of *zen*. *Proc. Natl Acad. Sci. USA* **93**, 8479-8484 (1996).

25 Brown, S. J. *et al.* Sequence of the *Tribolium castaneum* homeotic complex: The region corresponding to the *Drosophila melanogaster* Antennapedia Complex. *Genetics* **160**, 1067-1074 (2002).

26 van der Zee, M., Berns, N. & Roth, S. Distinct functions of the *Tribolium zerknüllt* genes in serosa specification and dorsal closure. *Curr. Biol.* **15**, 624-636 (2005).

27 Angelini, D. R. & Jockusch, E. L. Relationships among pest flour beetles of the genus *Tribolium* (Tenebrionidae) inferred from multiple molecular markers. *Mol. Phylogenet. Evol.* **46**, 127-141 (2008).

28 Svobodova, E., Kubikova, J. & Svoboda, P. Production of small RNAs by mammalian Dicer. *Pflugers Archiv : European journal of physiology* **468**, 1089-1102 (2016).

29 Benton, M. A., Akam, M. & Pavlopoulos, A. Cell and tissue dynamics during *Tribolium castaneum* embryogenesis revealed by versatile fluorescence labeling approaches. *Development* **140**, 3210-3220 (2013).

30 Handel, K., Grünfelder, C. G., Roth, S. & Sander, K. *Tribolium* embryogenesis: a SEM study of cell shapes and movements from blastoderm to serosal closure. *Dev. Genes Evol.* **210**, 167-179 (2000).

31 Koelzer, S., Kölsch, Y. & Panfilio, K. A. Visualizing late insect embryogenesis: Extraembryonic and mesodermal enhancer trap expression in the beetle *Tribolium castaneum*. *PLoS One* **9**, e103967 (2014).

32 Truckenbrodt, W. The embryonic covers during blastokinesis and dorsal closure of the normal and of the actinomycin D treated egg of *Odontotermes badius* (Hav.) (Insecta, Isoptera). *Zoologische Jahrbücher (Abteilung für Anatomie und Ontogenie der Tiere)* **101**, 7-18 (1979).

33 Cohen, E., Allen, S. R., Sawyer, J. K. & Fox, D. T. Fizzy-related dictates a cell cycle switch during organ repair and tissue growth responses in the Drosophila hindgut. *eLife* **7**, e38327 (2018).

34 Schaeffer, V., Althauser, C., Shcherbata, H. R., Deng, W. M. & Ruohola-Baker, H. Notch-dependent Fizzy-related/Hec1/Cdh1 expression is required for the mitotic-to-endocycle transition in Drosophila follicle cells. *Curr. Biol.* **14**, 630-636 (2004).

35 Horn, T. & Panfilio, K. A. Novel functions for *Dorsocross* in epithelial morphogenesis in the beetle *Tribolium castaneum*. *Development* **143**, 3002-3011 (2016).

36 Dearden, P., Grbic, M., Falciani, F. & Akam, M. Maternal expression and early zygotic regulation of the *Hox3/zen* gene in the grasshopper *Schistocerca gregaria*. *Evol. Dev.* **2**, 261-270 (2000).

37 Banks, C. J., Joshi, A. & Michoel, T. Functional transcription factor target discovery via compendia of binding and expression profiles. *Sci. Rep.* **6**, 20649 (2016).

38 Shah, N., Dorer, D. R., Moriyama, E. N. & Christensen, A. C. Evolution of a large, conserved, and syntenic gene family in insects. *G3 (Bethesda)* **2**, 313-319 (2012).

39 Strobl, F., Anderl, A. & Stelzer, E. H. A universal vector concept for a direct genotyping of transgenic organisms and a systematic creation of homozygous lines. *eLife* **7**, e31677 (2018).

40 Liu, Q. *et al.* Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development. *eLife* **7**, e34594 (2018).

41 Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S. & Aggarwal, A. K. Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature* **397**, 714-719 (1999).

42 Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151-1155 (2000).

43 Crocker, J. *et al.* Low affinity binding site clusters confer Hox specificity and regulatory robustness. *Cell* **160**, 191-203 (2015).

44 Porcelli, D., Fischer, B., Russell, S. & White, R. Chromatin accessibility plays a key role in selective targeting of Hox proteins. *Genome Biol.* **20**, 115 (2019).

45 Davidson, E. H. & Erwin, D. H. Gene regulatory networks and the evolution of animal body plans. *Science (New York, N.Y.)* **311**, 796-800 (2006).

46 Hart, Y. & Alon, U. The utility of paradoxical components in biological circuits. *Molecular Cell* **49**, 213-221 (2013).

47 Kumar, R. *et al.* Role of Homothorax in region specific regulation of Deformed in embryonic neuroblasts. *Mech. Devel.* **138 Pt 2**, 190-197 (2015).

48 Streit, A. *et al.* Conserved regulation of the Caenorhabditis elegans labial/Hox1 gene ceh-13. *Dev. Biol.* **242**, 96-108 (2002).

49 Schoppmeier, M., Fischer, S., Schmitt-Engel, C., Löhr, U. & Klingler, M. An ancient anterior patterning system promotes *caudal* repression and head formation in Ecdysozoa. *Curr. Biol.* **19**, 1811-1815 (2009).

50 Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450-461 (2007).

51 Gavin-Smyth, J., Wang, Y. C., Butler, I. & Ferguson, E. L. A genetic network conferring canalization to a bistable patterning system in Drosophila. *Curr. Biol.* **23**, 2296-2302 (2013).

52 Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298**, 1241-1245 (2002).

53 Nunes da Fonseca, R. *et al.* Self-regulatory circuits in dorsoventral axis formation of the short-germ beetle *Tribolium castaneum*. *Dev. Cell* **14**, 605-615 (2008).

54 Schmidt-Ott, U., Rafiqi, A. M. & Lemke, S. in *Hox Genes: Studies from the 20th to the 21st Century* Vol. 689 *Advances in Experimental Medicine and Biology* (ed Jean S. Deutsch) 133-144 (Landes Bioscience, 2010).

55 Rafiqi, A. M., Lemke, S. & Schmidt-Ott, U. Postgastrular *zen* expression is required to develop distinct amniotic and serosal epithelia in the scuttle fly *Megaselia*. *Dev. Biol.* **341**, 282-290 (2010).

56 Reim, I., Lee, H.-H. & Frasch, M. The T-box encoding Dorsocross genes function in amnioserosa development and the patterning of the dorsolateral germ band downstream of Dpp. *Development* **130**, 3187-3204 (2003).

57 Panfilio, K. A. & Roth, S. Epithelial reorganization events during late extraembryonic development in a hemimetabolous insect. *Dev. Biol.* **340**, 100-115 (2010).

58 Orr-Weaver, T. L. When bigger is better: the role of polyploidy in organogenesis. *Trends Genet.* **31**, 307-315 (2015).

59 Hughes, C. L., Liu, P. Z. & Kaufman, T. C. Expression patterns of the rogue Hox genes *Hox3/zen* and *fushi tarazu* in the apterygote insect *Thermobia domestica*. *Evol. Dev.* **6**, 393-401 (2004).

60 Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531-1545 (1999).

61 Schmidt-Ott, U. & Kwan, C. W. Morphogenetic functions of extraembryonic membranes in insects. *Curr. Op. Insect Sci.* **13**, 86-92 (2016).

62 Miles, W. O. *et al.* Medea SUMOylation restricts the signaling range of the Dpp morphogen in the Drosophila embryo. *Genes Dev.* **22**, 2578-2590 (2008).

63 Podos, S. D., Hanson, K. K., Wang, Y.-C. & Ferguson, E. L. The DSmurf Ubiquitin-protein ligase restricts BMP signaling spatially and temporally during *Drosophila* embryogenesis. *Dev. Cell* **1**, 567-578 (2001).

64 Tvrdik, P. & Capecchi, M. R. Reversal of Hox1 gene subfunctionalization in the mouse. *Dev. Cell* **11**, 239-250 (2006).

65 Panfilio, K. A. Late extraembryonic development and its *zen-RNAi*-induced failure in the milkweed bug *Oncopeltus fasciatus*. *Dev. Biol.* **333**, 297-311 (2009).

66 Brown, S. J. *et al.* The red flour beetle, *Tribolium castaneum* (Coleoptera): a model for studies of development and pest biology. *Cold Spring Harb Protoc* **2009**, pdb emo126 (2009).

67 Herndon, N. *et al.* Enhanced genome assembly and a new official gene set for *Tribolium castaneum* - from a draft to a reference genome. *BMC Genomics* **21**, 47 (2020).

68 Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nuc. Acids Res.* **25**, 3389-3402 (1997).

69 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

70 Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* **32**, W273-279 (2004).

71 Mayor, C. *et al.* VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**, 1046-1047 (2000).

72 Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist *Nucleic Acids Res.* **36(Web Server issue)**, W465-W469 (2008).

73 Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).

74 Stothard, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**, 1102, 1104 (2000).

75 R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. *Vienna, Austria* **URL** https://www.R-project.org/ (2016).

76 Cribari-Neto, F. & Zeileis, A. Beta Regression in R. *J Stat Softw* **34**, 1-24 (2010).

77 Mackrodt, D. *Etablierung und Funktion maternaler Proteingradienten im Tribolium Blastoderm* Ph.D. (Dr. rer. nat.) thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, (2016).

78 Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* (ed John M. Walker) 571-607 (Humana Press, 2005).
79 Andrews, S. FASTQC. A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ **version 0.11.5, released 8 March 2016** (2010).
80 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
81 Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A. & Dewey, C. N. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**, 493-500 (2010).
82 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
83 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
84 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
85 Busby, M. A., Stewart, C., Miller, C. A., Grzeda, K. R. & Marth, G. T. Scotty: a web tool for designing RNA-Seq experiments to measure differential gene expression. *Bioinformatics* **29**, 656-657 (2013).
86 Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676 (2005).

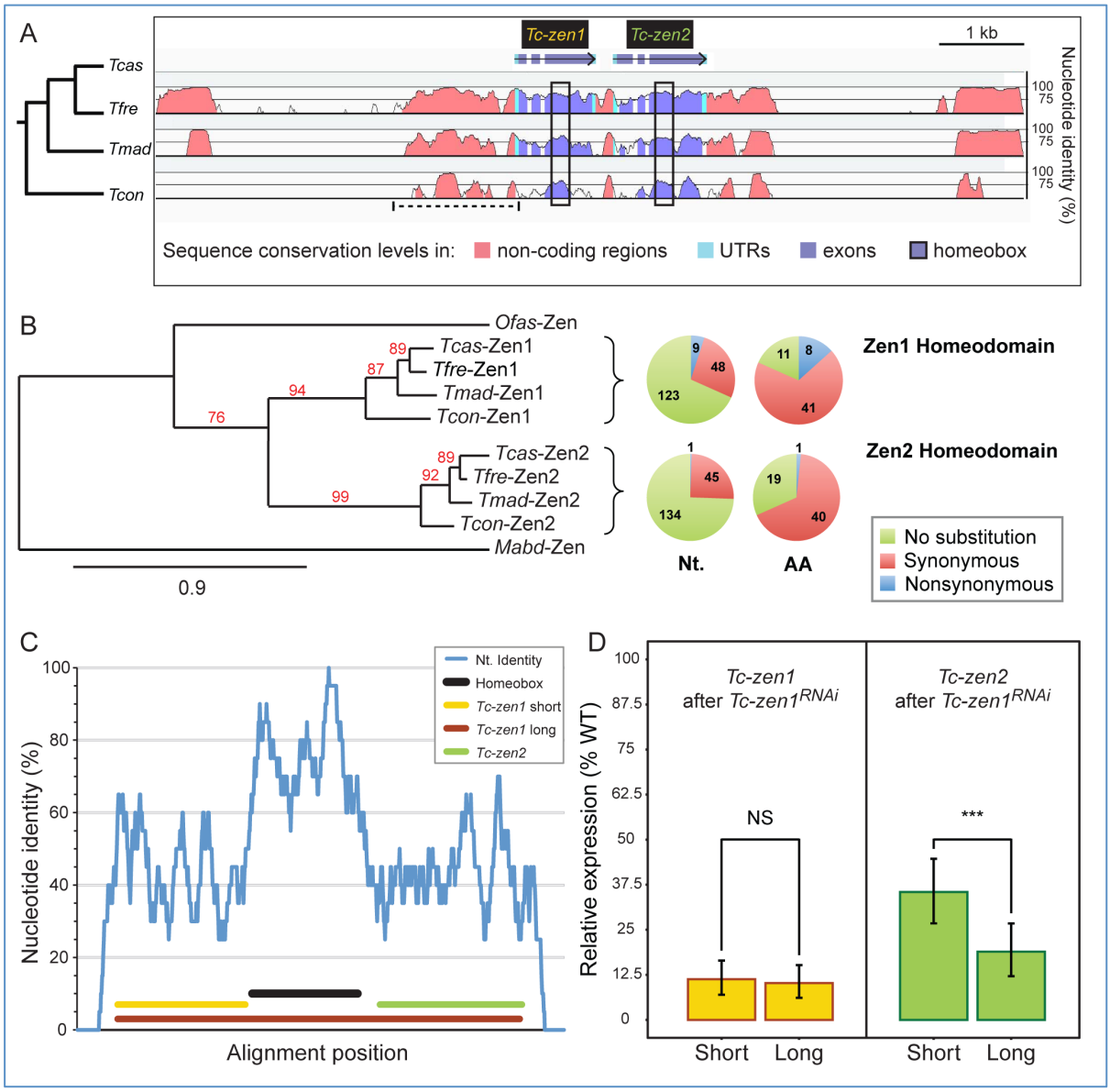

Fig 1. High conservation of *Tribolium zen* orthologues and paralogues.

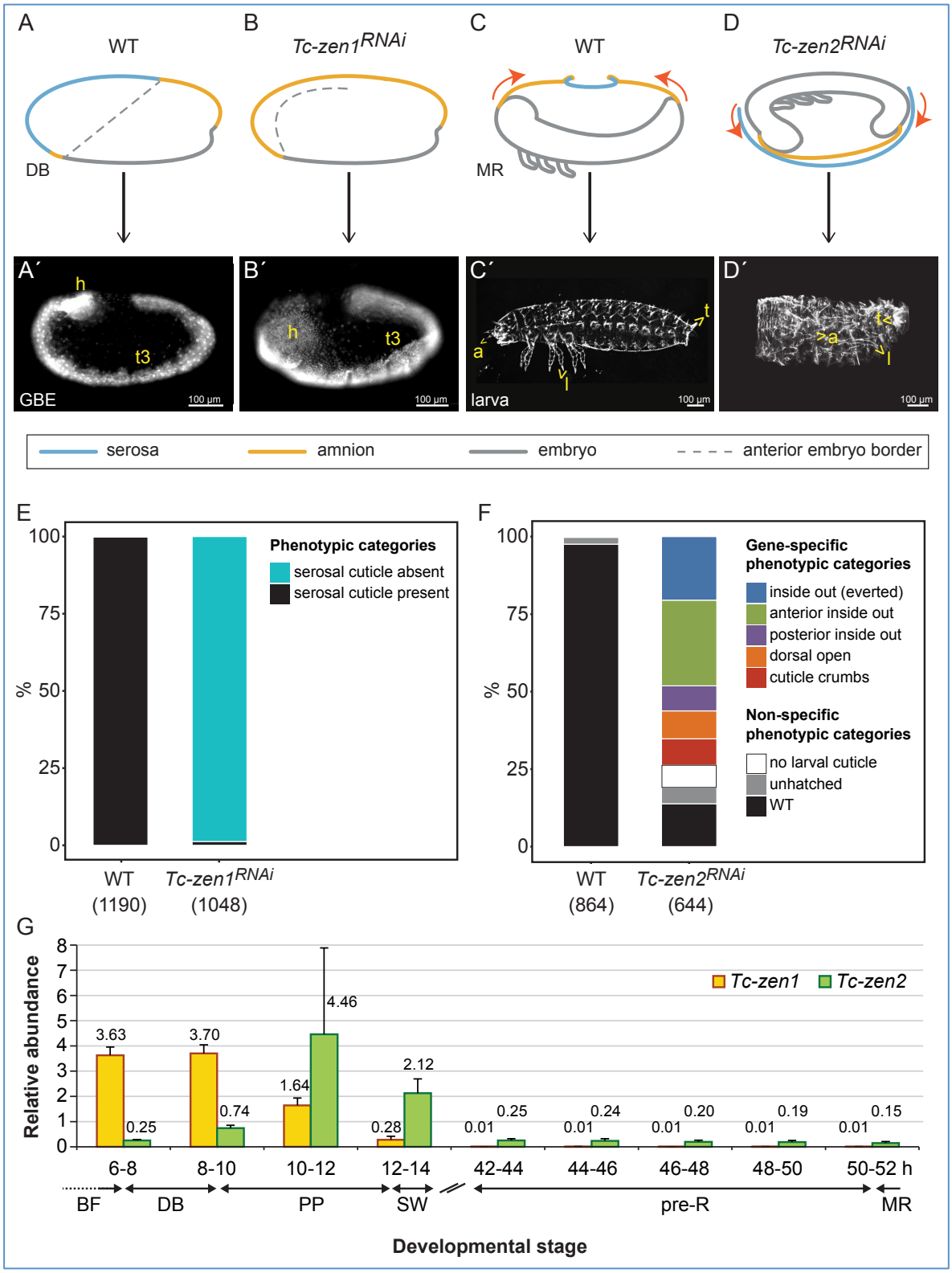

Fig 2. *Tc-zen* paralogue roles in early specification (*Tc-zen1*) or late EEM morphogenesis (*Tc-zen2*).

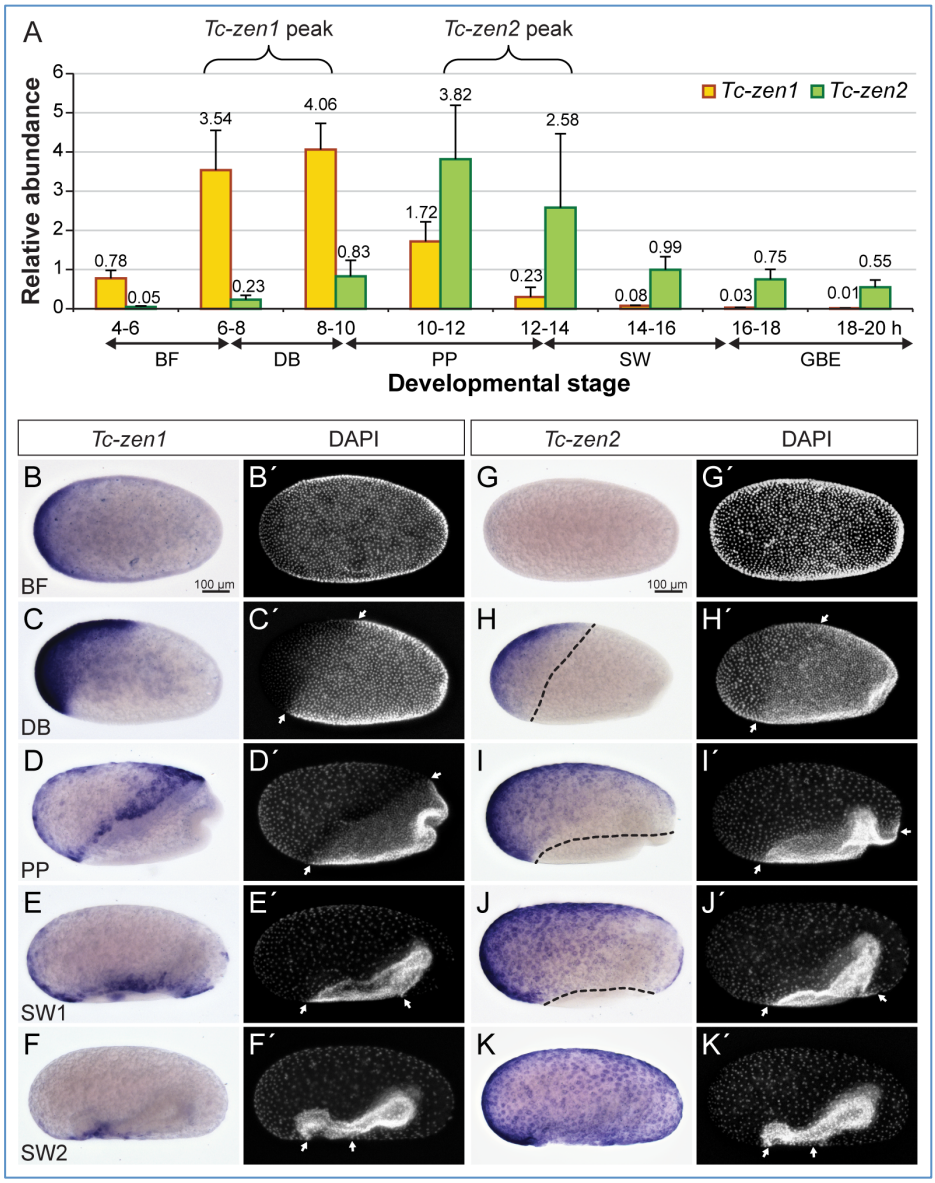

Fig 3. Transcript expression dynamics of the *Tc-zen* paralogues during early embryogenesis.

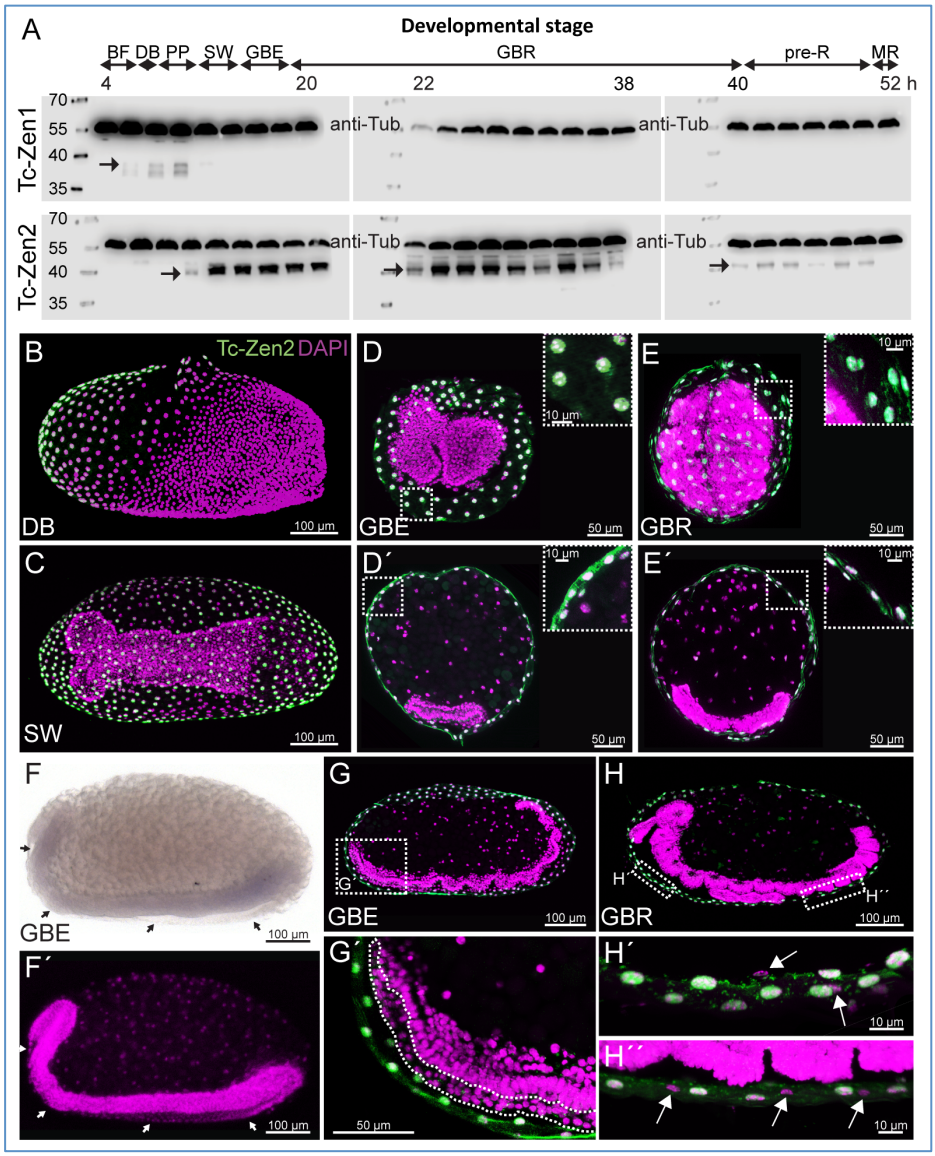

Fig 4. Paralogue-specific Tc-Zen protein expression time courses and spatial restriction of Tc-Zen2.

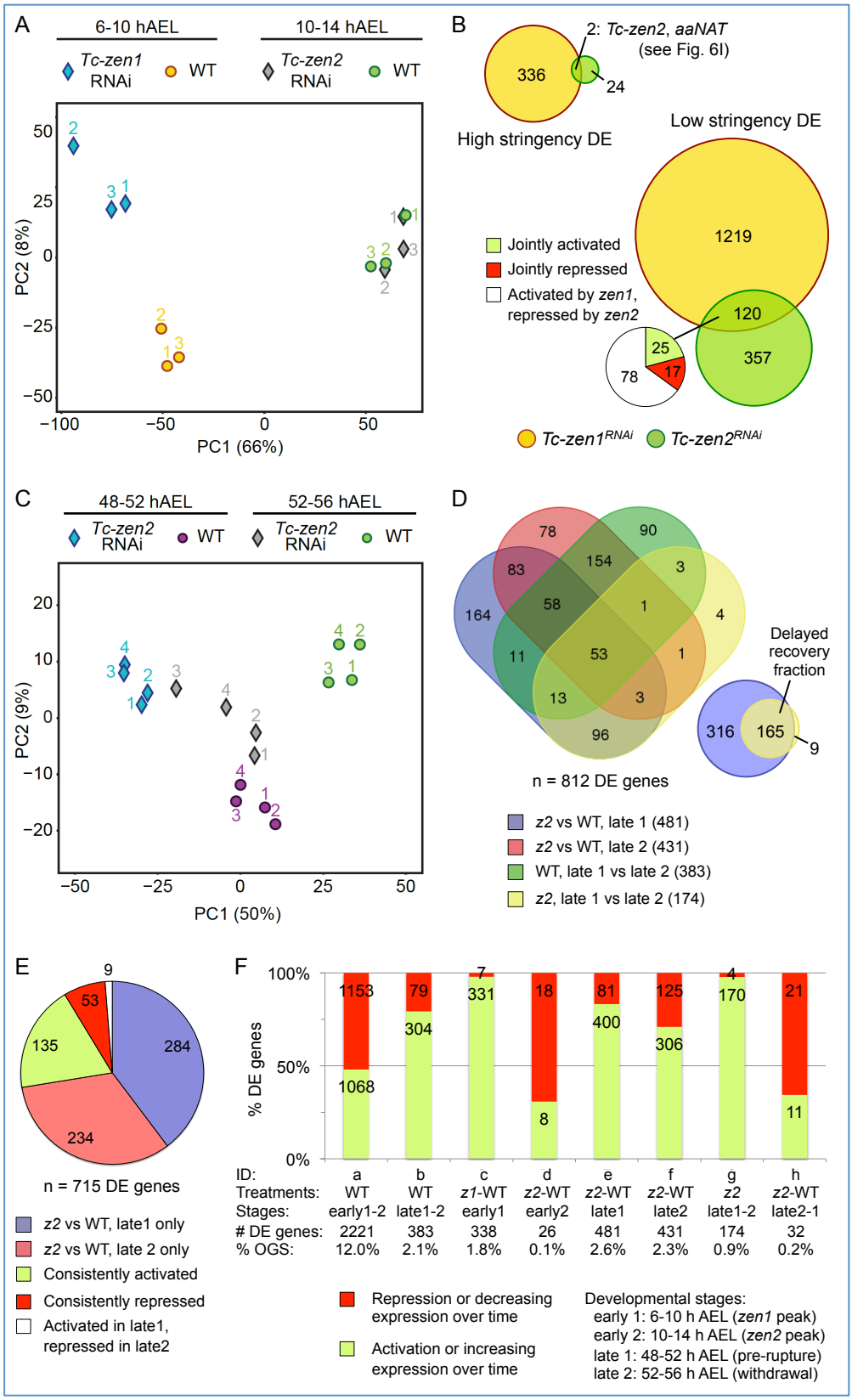

Figure 5. Tc-zen genes' transcriptional control: relative impact, overlap, and stage-specific dynamics revealed by RNA-seq after RNAi.

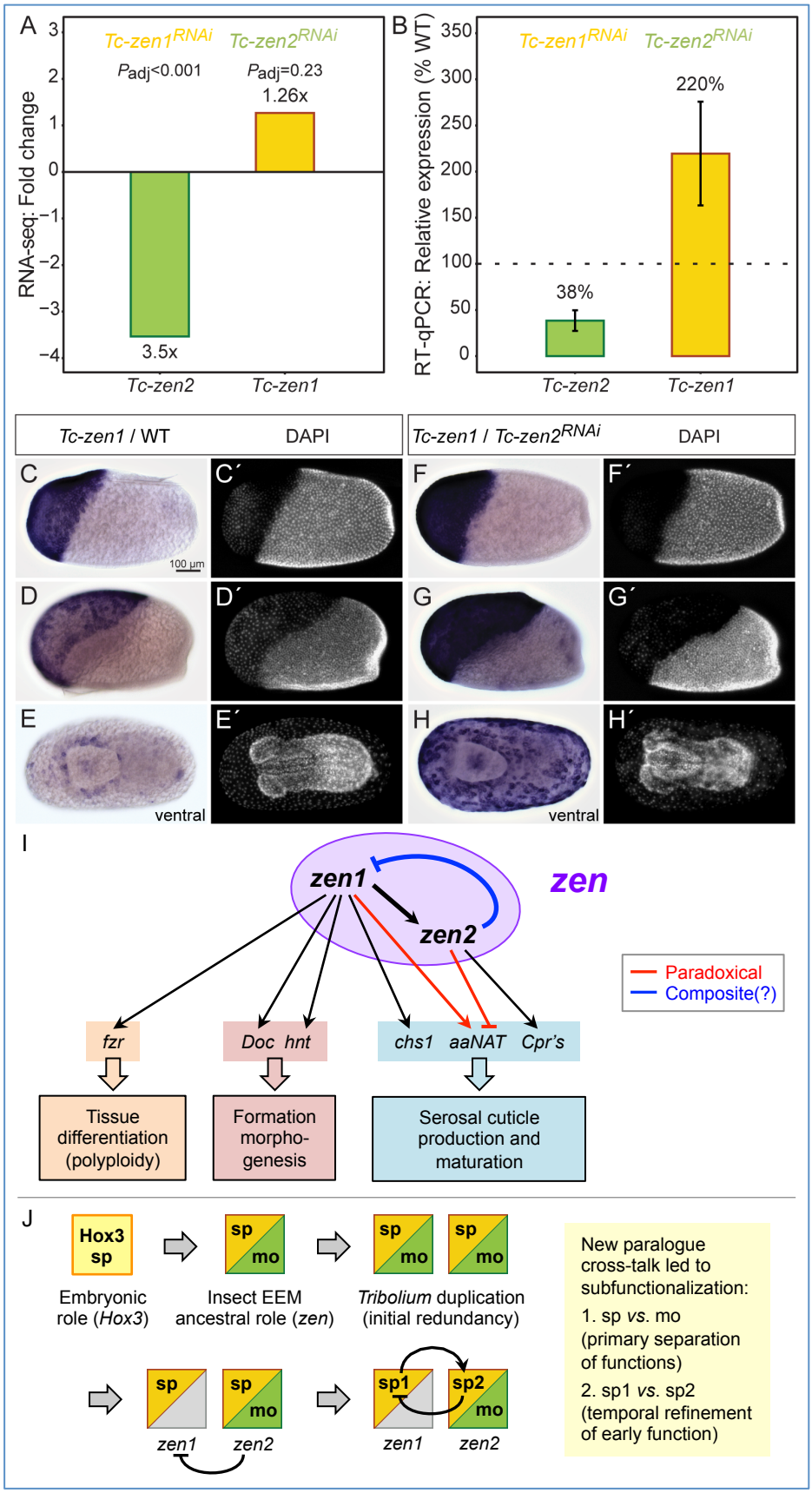

Fig 6. Evidence and implications of a *zen* paralogue regulatory module.