Ankyrin domain encoding genes resulting from an ancient horizontal transfer are functionally integrated into developmental gene regulatory networks in the wasp *Nasonia*

Daniel Pers[1] and Jeremy A. Lynch[&,1]

[1]Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL, USA

[&]Author for correspondence

Jeremy A. Lynch
University of Illinois at Chicago
MBRB 4020
900 S. Ashland Ave
Chicago, IL 60607
312-996-5460
jlynch42@uic.edu

ABSTRACT

Background

How and why regulatory networks incorporate additional components, and how novel genes are maintained and functionally integrated into developmental processes are two important and intertwined questions whose answers have major implications for the evolution of development. We recently described a set of novel genes with robust and unique expression patterns along the dorsal-ventral axis of the embryo of the wasp *Nasonia*. Given the unique evolutionary history of these genes, and their apparent integration in to the dorsal-ventral (DV) patterning network, they are collectively an excellent model to study the evolution of regulatory networks, and the fates of novel genes.

Results

We have found that the novel DV genes are part of a large family of rapidly duplicating and diverging ankyrin domain encoding genes that originated most likely by horizontal transfer from *Wolbachia* in a common ancestor of the wasp superfamilly Chalcidoidea. We tested the function of those ankyrin encoding genes expressed along the DV axis and found that they participate in early embryonic DV patterning. We also developed a new wasp model system (*Melittobia*) and found that some functional integration of ankyrin genes have been preserved for over 90 million years, while others are lineage specific.

Conclusions

Our results indicate that regulatory networks can incorporate novel genes that then become necessary for stable and repeatable outputs. Even modest role in

69 developmental networks may be enough to allow novel or duplicate genes to be
70 maintained in the genome and become fully integrated network components.
71
72 **KEYWORDS**
73 Regulatory networks, Development, EvoDevo, Ankyrin, Chalcidoidea, *Nasonia*,
74 dorsoventral patterning, embryo, *Wolbachia,* gene duplication, Horizontal Gene Transfer
75 **Background**
76 Gene regulatory networks (GRNs) coordinate the expression of mRNA and
77 proteins in a spatiotemporal manner to bring about a specific developmental output [1].
78 The complex webs of interacting nodes and modules that make up GRNs are vital for
79 establishing patterning, morphogenesis, and ultimately an organism's body plan [2].
80 Perturbations to these networks should result in novel developmental outputs. However,
81 canalization and developmental redundancy can conceal underlying genetic variation
82 and phenotypic plasticity. This quality enables GRNs to weather large variations in
83 genomic and environmental inputs, without disrupting the phenotypic output of the
84 network [3-6].
85 These properties of GRNs raise questions about how developmental
86 mechanisms can evolve. Since robust networks can absorb large genetic changes
87 without causing major changes in developmental output, it seems that a large threshold
88 must be overcome in order to achieve a new phenotype [7]. Thus, robustness
89 paradoxically may make GRNs less able to respond to evolutionary pressures, since
90 most mutations will not produce phenotypes visible to natural selection. Thus, we might
91 expect robust developmental GRNs to be mostly static in evolutionary time in the

absence of major phenotypic change. However, there are many well known examples where developmental processes have been apparently unchanged, while the molecular basis of development is highly diverged [8-10]

Whether these changes are fixed because they provide some selectable improvement on the developmental process of interest, are indirect responses to selection on modules that are reused in other developmental processes, or are random is not well characterized. The development of methods to circumvent the candidate gene approach in a very wide variety of species facilitates comprehensive characterization of developmental GRNs at high phylogenetic resolution. This can allow hypotheses about the evolution of development to be tested robustly, and will lead to a deep understanding of the pattern and process of GRN evolution.

The GRN that patterns the embryonic dorsoventral (DV) axis of the wasp *Nasonia vitripennis* has been shown to be a good model to study novelty and the evolution of gene networks. Having split from *Drosophila melanogaster* over 300 MYA [11], the two have converged on a similar mode of embryogenesis [12], and share a nearly identical expression of tissue-specific marker genes just prior to gastrulation [13]. We have previously shown that most genes differentially expressed along the DV axis of the *Nasonia* embryo are not conserved components of the *Drosophila* DV GRN, making the comparison between the fly and wasp DV GRNs an ideal system for understanding how GRNs evolve while producing similar patterning results [14, 15].

A particularly interesting case of *Nasonia* specific DV GRN components are a set of 15 ankyrin domain containing genes, which do not have clear orthologs in *Drosophila* or in any other insects outside of the Superfamily Chalcidoidea. In fact, there is

115 evidence that these genes entered the genome of the ancestor of *Nasonia* through at
116 least one horizontal gene transfer (HGT) event, followed by several waves of duplication
117 and divergence. We have previously shown that these ankyrin domain encoding genes
118 are expressed in specific patterns along the DV axis [15], and here we demonstrate that
119 they also are functionally incorporated into the DV patterning GRN, and their loss leads
120 to variable disruptions to patterning. Through examination of another wasp, *Melittobia*
121 *digitata*, we also show that some of the functional incorporation is ancient within the
122 Superfamily, while there is also strong evidence of recent gains and/or losses of
123 function in the *Nasonia* and *Melittobia* lineages.
124 We propose that the properties of ankyrin domain containing proteins allow them
125 to rapidly gain interaction partners, and potentially adaptive functions in developmental
126 networks, which increases the likelihood that genes of this type will be maintained and
127 sometimes multiply in the course of genome evolution.
128
129 **Results**
130 **Identification of new families of apparently horizontally transferred, ankyrin**
131 **domain encoding, genes**
132 In our previous study we identified fifteen transcripts encoding ankyrin domain
133 proteins that appeared to be significantly regulated by the Toll and/or BMP signaling
134 pathways in the *Nasonia* embryo [15]. Further analysis of their expression showed that
135 6 of these genes are expressed laterally, 3 are expressed on the dorsal surface of the
136 embryo, one is expressed over the ventral midline, one has a complex pattern involving

late expression in dorsal tissues, and 4 with no clear differential expression along the DV axis (described in more detail below).

Our previous analysis indicated that four of these 15 genes possess a PRANC (**P**ox proteins **R**epeats of **AN**kyrin, **C**-terminal) domain at their C-termini. PRANC domains were originally described in Pox viruses. They were first described in a eukaryotic system upon the publication of the *Nasonia* genome, where a set of PRANC domain encoding genes were found to be integrated into the genome, and which are highly similar PRANC domain proteins in their endosymbiotic bacteria, *Wolbachia*[16, 17]. The similarity (both in the PRANC domains, and their association with ankyrin repeats) of the PRANC encoding genes integrated into the *Nasonia* genome to those found in the *Wolbachia* genome led to the hypothesis that the *Nasonia* PRANC genes originated from a HGT from *Wolbachia*.

While the remaining 11 DV regulated ankyrin domain encoding genes do not appear to have PRANC domains, we believe that they entered the *Nasonia* genome through similar processes of horizontal transfer, gene duplication and rapid molecular divergence.

If we focus on BLASTp searches using any of the 15 *Nasonia* DV ankyrin domain genes as queries against well annotated genomes from *D. melanogaster* or *Apis meliferra* the results are invariably "canonical" ankyrin genes (i.e., *ankyrin-1* or *ankyrin-2*), that are highly conserved in all insect species. As expected, *Nasonia* possesses clear direct orthologs to such canonical ankyrin genes that are highly conserved at the amino acid level and these gene families are clearly distinct from the genes we propose are horizontally transferred (Additional File 1 and 2).

Additional evidence that these genes originated outside of the normal course of vertical transmission from generation to generation is the pattern of unrestricted BLASTp results and the phylogenetic relationships of the top 100 protein BLAST hits for each *Nasonia* ankyrin (Additional File 3) which do not follow expected phylogenetic patterns (i.e., stronger hits throughout the hymenoptera, then hits in Diptera, Coleoptera, Lepidoptera, other insects, etc...), that are observed for the canonical ankyrin genes.

As would be expected, the top hits for most of the 15 *Nasonia* sequences come from *Trichomalopsis* (a sister genus to *Nasonia*), except one where there is a recent *Nasonia* paralog almost identical to it (Additional File 3). This indicates that the ancestors of the 15 DV regulated ankyrin genes were present in the common ancestor of the *Nasonia* and *Trichomalopsis* genera. In addition, sequences from the fig wasp *Ceratosolen solmsi* [18], Copidosoma floridanum and *Trichogramma pretiosum* [19] (representing the Families Agaonidea, Encyrtidae, and Trichogrammitidae within the Superfamily Chalcidoidea, respectively) are among the top hits for most of the *Nasonia* proteins, and cluster near to the *Nasonia* query sequence in the phylogenies (Additional File 3).

Outside of the chalcid wasps, strong hits are scattered across the tree of life. Taxa with proteins showing strong similarity to each *Nasonia* DV ankyrin gene include the ant *Pseudomyrmex gracilis*, the single celled eukaryote *Trichomonas vaginalis*, the bee *Ceratina capitata*, the sea urchin *Strongylocentrotus purpuratus*, and the amoeba endosymbiont *Candidatus Amoebophilus asiaticus* [20]. Other taxa occur with less frequency, including *Wolbachia* sequences (Additional File 3). Importantly, canonical

ankyrins from *Nasonia* or other insect species do not appear among the top hundreds of hits and do not cluster with our genes of interest in phylogenies.

We attempted to phylogenetically analyze the Nasonia DV-ankyrin genes in the context of their best blast hits in other animals and microbes to see if they cluster with prokaryotic or canonical eukaryotic ankyrin sequences. However, we found that aligning our relatively small, rapidly evolving proteins with proteins encoding varying numbers of ankyrin domains of varying conservation made phylogenetic analyses difficult, and in our hands uninformative (not shown).

Instead, we decided to focus on the regions C-terminal to the ankyrin repeats in the *Nasonia* DV ankyrin domain genes. These regions range from ~100-200 amino acids, except in two sequences (-D and -E), which lack C-terminal sequence beyond the ankyrin domains. We knew that this region was predicted to contain a PRANC domain in four of our DV ankyrin proteins, but we did not know how similar to sequences outside of the wasps it would be. We also hypothesized that the remaining C-termini maintained cryptic similarity to the ancestral PRANC domain.

Because the C-termini of our DV ankyrin domain genes seem to be less well conserved, we used the more sensitive, iterative PSI-BLAST [21] approach to identify similar sequences in the NCBI nonredundant (nr) database. We used default parameters, including only using genes that were above the threshold in round one as templates to generate the pattern for the second-round search. We then took all of the aligning sequences that were above the significance threshold given by PSI-BLAST and subjected them to phylogenetic analysis. We only discuss genes that were above the threshold in the second round (with the exception of Nv-CLANK-L, which required 4

rounds), as this was sufficient to identify the first non-insect and microbial sequences. DV ankyrins -D, and -E were not analyzed as they lack sequence C-terminal to the ankyrin domains.

The taxa that appear in these queries are much more limited than what was found using the full DV ankyrin protein sequences as queries, indicating that complexity of aligning the constrained and repetitive ankyrin domains can indeed give spurious signals of homology. The close relationship of *Nasonia* DV ankyrins and other orphan ankyrin domain encoding genes in Chalcidoidea is accentuated, as multiple sequences from *Ceratosolen*, *Copidosoma*, and *Trichogramma* cluster more robustly with large numbers of *Nasonia* sequences in this analysis (Figure 1A, for example, Additional File 4). We also find sequences from species that also showed up strongly when using the full protein as a query, particularly from the ant *Pseudomyrmex* and bee *Ceratina*. We also find a large number of hits from the Braconid wasp *Microplitus demolitor* and the whitefly *Bemisia tabaci* in all PSI-BLASTs. The sequences from these insects cluster closely with others from the same species, and also with particular *Wolbachia* sequences (Figure 1B, for example, Additional File 4), again indicating recent HGT in these organisms.

Outside of these insect hits were primarily from prokaryotes and some poxviruses. *Wolbachia* species had the strongest and most common prokaryotic hits, but many *Orientia* (a Rickettsial intracellular parasite where PRANC domains were also found in the *Nasonia* genome publication [17]) sequences were also found (Additional File 4). The e-values of Wolbachia hits ranged from E-05 to E-60 (Fig. 1C, for example). Importantly, all of the hits aligned to the C-terminal region of small *Wolbachia*

proteins that also have ankyrin domains toward the N-terminus. Some of the hits in these analyses are annotated as PRANC domains, but the majority of them are not, despite highly significant similarity.

Overall, these results strongly indicate that the ankyrin domain containing genes that were identified in the PSI-BLAST analyses have a common history, despite the fact that they occur in a scattered phylogenetic distribution in a few insect lineages, intracellular bacteria, and viruses. We find it extremely unlikely that these proteins with conserved C-terminal, PRANC-like motifs that are directly downstream from relatively well conserved ankyrin domains, would have evolved convergently by chance multiple times.

Rather, we believe that the pattern we uncovered indicates multiple instances of HGT: At least four recent ones in lineages leading to *Pseudomyrmex*, *Bemisia,* *Ceratina*, and *Microplitus*, and one ancient one in a common ancestor of the superfamily Chalcidoide (around 150 million years ago [22]). We propose to name this later family of proteins **C**halcidoidea **L**ineage specific **ANK**yrin domain encoding genes (***CLANK***s). We will henceforth discuss the DV ankyrin domain proteins as *Nasonia vitripennis* CLANKs (*Nv*-CLANK) -A through -O. The relationships between our CLANK nomenclature and gene identification numbers in different annotations are given in Additional File 5.

While we strongly favor the hypothesis that the CLANKs entered the genomes of Chalcid wasps through HGT based on the evidence given above, the complex relationships and genetic exchange back and forth among prokaryotes, viruses and eukaryotes [16, 20, 23], make proving this idea beyond a shadow of a doubt a daunting

task, well beyond the scope of this manuscript. Whatever the case, CLANKs are new genes in the wasps relative to the rest of the insects, and we would like to understand why they have been maintained over the course of more than 150 million years of evolution in this clade.

Characterization of protein and genomic sequence features

Characterization of the amino acid sequences and genomic context of these CLANK genes supports that this family of genes likely has a long history, and has gone extensive sequence change in the course of their evolution. These 15 genes code for 18 proteins that vary in size from 255 to 717 amino acids. Most of the genes encoding *Nv-CLANKs* are interrupted by at least one intron, and they are found spread across the 5 chromosomes of *Nasonia* (Fig. 2). Sequence alignment of these 15 protein sequences yielded very few residues conserved across the CLANKs [24] (data not shown). Phylogenetic analysis [25] of the DV CLANKs gives two major clades (Fig. 2). Domain analysis with Interproscan [26] confirmed the four CLANKS have officially annotated PRANC domains (*Nv*-CLANK-F, -J, -N, -O), and mapping these on the phylogeny shows that this domain has likely been lost or degraded beyond recognition in many lineages (Fig. 2). This analysis also showed that the number of Ankyrin repeats within a protein varies from 6 to 13 throughout the length of the proteins, except for the last ~100 amino acids, where the PRANC-like domains reside.

Overall, there appears to be no correlations between these genetic features and regulation, expression patterns, or phylogenetic relatedness (Fig. 2), except for the two pairs of recent paralogy (*Nv*-CLANK-F and -N, and -A and -H).

Detailed characterization of DV CLANK embryonic expression

RNA expression patterns for the 15 CLANK genes have previously been mentioned [15], but not fully described. Thus, *in situ* hybridization experiments were thus repeated and analyzed in more detail over a longer developmental time frame, and transcripts were grouped according to their expression patterns. Four CLANKS (*-B, -C, -D, -J*) have no patterned expression at any time in embryogenesis (Additional File 6) and will not be discussed further.

Laterally expressed CLANKs

Six of the fifteen *Nv-CLANK* transcripts are expressed in a lateral domain at one or more time points during embryogenesis; however, this expression is quite dynamic. Three *Nv-CLANKs* (*-G*, *-H*, *-K*) show an overall expansion of expression, while three *Nv-CLANKs* (*-A*, *-E*, *-I*) are characterized by a shrinking of their expression domain.

Nv-CLANK-G is ubiquitously expressed during the pre-blastoderm and early blastoderm stages of development (Fig. 3A1-32). As the blastoderm undergoes additional rounds of division and begins to cellularize, *Nv-CLANK-G* is expressed first as a band encircling the anterior end of the embryo (Fig. 3A3) and then expands posteriorly creating a gradient with highest levels in that initial anterior domain (Fig. 3A4). Expansion is missing from both poles, and also from the dorsal midline of the embryo (Fig. 3A5). During gastrulation, expression is restricted to a segmental pattern in the cells that give rise to the central nervous system (CNS) (Fig. 3A6).

The expression pattern of *Nv-CLANK-H* is very similar to *Nv-CLANK-G* during the middle and late blastoderm stages. It also is initially expressed in an anterior band that eventually forms a gradient of expression along the AP axis (Fig. 3B1-2). Again,

298 expression is absent at the AP poles and along the dorsal midline. However, while *Nv-*
299 *CLANK-G* was expressed maternally, *Nv-CLANK-H* has no early expression (data not
300 shown) and is ubiquitously expressed at very low levels during gastrulation instead of
301 being localized to CNS precursors (Fig. 3B3).

302 *Nv-CLANK-K* is expressed maternally and ubiquitously at low levels (Fig. 3C1)
303 before being expressed in an anterior-lateral domain with expression lacking at both the
304 ventral and dorsal midline and strongest near the anterior pole (Fig. 3C2). This domain
305 then shifts and expands posteriorly into a more evenly expressed lateral domain, with
306 inhibition of expression ventrally and at the poles (Fig. 3C3). During gastrulation
307 expression is lost completely (data not shown).

308 *Nv-CLANK-A* is initially expressed ubiquitously at very low levels before becoming
309 localized in a broad lateral domain in the early blastoderm (Fig. 4A1-A2). The lateral
310 domain then shrinks into two discrete bands in the trunk of the embryo, before
311 expanding first dorsoventrally and then anteroposteriorly into one lateral band (Fig. 4A3-
312 A5). At times, expression is missing along the ventral and dorsal midlines; however, this
313 is variable and dynamically changes as the blastoderm undergoes further division and
314 cellularizes. Expression is lacking in the gastrulating embryo (Fig. 4A6).

315 *Nv-CLANK-E* expression is lacking or at very low levels both before and after the
316 blastoderm stages (data not shown). Expression is in a broad band from the nearly the
317 middle of the embryo to just anterior of the posterior pole during the syncytial
318 blastoderm (Fig. 4B1). During cellularization, this band retracts to a smaller, weaker,
319 expression domain in the posterior of the embryo (Fig. 4B2-B3).

The expression pattern of *Nv-CLANK-I* in the mid to late blastoderm also resembles that of *Nv-CLANK-G* and *-H*. Again, there is a lateral expression domain with expression missing from the two poles and the dorsal midline (Fig. 4C1); however, the dynamics of this transcript differ. Instead of expression first appearing in the anterior region and then appearing progressively towards the posterior pole, *Nv-CLANK-I* is initially expressed in this broad domain, spanning the trunk of the embryo, having slightly higher expression in the anterior and posterior regions. During gastrulation, the anterior expression is lost until there is just a posterior expression band (Fig. 4C2) and then a posterior spot (Fig. 4C3). *Nv-CLANK-I* is not expressed maternally or in the early blastoderm. (data not shown).

Dorsally expressed CLANKs

Three of the fifteen transcripts are expressed dorsally during embryogenesis. *Nv-CLANK-N* and *-O* both lack maternal expression and are first expressed in the syncytial blastoderm. *Nv-CLANK-N* is expressed strongly at the anterior and posterior poles and appears to have weak and variable expression along the dorsal midline (Fig. 5A1-A3). *Nv-CLANK-O* is expressed evenly along the dorsal midline from pole to pole, stably expressed throughout syncytial divisions and cellularization (Fig. 5B1-B2). While *Nv-CLANK-N* lacks expression during gastrulation, *Nv-CLANK-O* is expressed dorsally, surrounding the extraembryonic material (Fig. 5B3).

Nv-CLANK-F is initially expressed at a low ubiquitous level (Fig. 5C1) before gaining expression at the anterior and posterior poles (Fig. 5C2). Expression then expands along the dorsal midline, anterior to posteriorly, ultimately connecting the two poles (Fig. 5C2-C4). In addition, there is dynamic ventral expansion that varies from embryo to

343 embryo and from stage to stage. In some cases, the dorsal stripe expands past the
344 anterior pole, onto the ventral half of the embryo, creating an anterior cap (Fig. 5C4).
345 This expansion is accompanied by a perpendicular band, encircling the posterior end of
346 the embryo's trunk (Fig. 5C5). In other cases, the expansion does not cross into the
347 ventral half, but still expands to broad domains at the two poles, while remaining narrow
348 in the mid-trunk region (Fig. 5C6). As the onset of gastrulation nears, this expansion,
349 and the initial dorsal stripe, retracts, leaving a strong patch of expression at the anterior
350 pole, a lighter and smaller domain at the posterior pole, and a weak stripe perpendicular
351 to the dorsal midline in the anterior-dorsal region of the embryo (Fig. 5C7). While the
352 anterior patch remains strong at the start of gastrulation, the other two expression
353 domains weaken quickly (Fig. 5C8). Differential expression is eventually lost during
354 gastrulation, and the whole embryo exhibits weak, ubiquitous expression (Fig. 5C9).

355 **Ventrally expressed CLANK**

356 *Nv-CLANK-M* is the only transcript in this family with ventral expression. It starts as a
357 narrow stripe in the early blastoderm, much like *Nv-twist* (Fig. 6A1). Also, like *Nv-twi,* it
358 broadens later in development (Fig. 6A2). However, it never takes on a the typical
359 "slug" shape of the presumptive mesoderm and begins to disappear as gastrulation
360 initiates. The pattern of disappearance roughly coincides with regions being covered by
361 the lateral ectoderm (Fig. 6A3).

362 **Ubiquitously and postgastrular expressed CLANK**

363 *Nv-CLANK-L* is strongly expressed both maternally (Fig. 6B1) and zygotically (Fig.
364 6B2-B3). Prior to gastrulation, expression is ubiquitous with the exception of there being
365 no detectable expression within the budding pole cells (Fig. 6B2). During gastrulation

366 there is a moderate level of expression throughout the embryo; however, there are
367 highly elevated levels of expression in the area that will become the head and in regions
368 just anterior and posterior to the extraembryonic material (Fig. 6B3).

369 **Reduction of CLANK transcripts results in significant increases in embryonic**
370 **lethality**

371 In order to understand the functional significance of this DV expression of the
372 CLANKs in *Nasonia*, parental RNA interference (pRNAi), where double stranded RNA is
373 injected into female pupae and its effects are examined in the embryos she produces
374 [27], was used to knockdown each of the 11 genes with detectable DV expression
375 patterns. We first analyzed the gross effects of knockdown on embryonic survival to
376 hatching first instar larvae. Average embryonic lethality of the 11 knockdowns ranged
377 from 0.87% to 12.19% of embryos plated (Fig. 7). In all cases the frequency of lethality
378 was larger than in control-injected pupae (0.65%). The difference in lethality was
379 statistically significant ($p < 0.05$) for 6 of the 11 transcripts tested (Fig. 7).

380 In the course of these experiments focusing on embryonic development, we
381 observed that the pupal injections also had significant and severe effects on successful
382 pupal development for Nv-CLANKs. The effect was particularly strong in *Nv-CLANK-E*,
383 *-K*, *-L* where a quarter or less of injected pupae completed metamorphosis, compared
384 the 60% rate for mock injected wasps. (Additional File 7). This suggests that these
385 transcripts may have additional functions in developmental or physiological processes
386 of pupation.

387 **CLANK transcripts levels are effectively reduced through pRNAi injections**

A major caveat of the pRNAi is that the degree by which a transcript is knocked down and the rate at which the system is turned over can vary from gene to gene. In order to test the effectiveness of each designed dsRNA, we quantified the knockdown using qPCR. Ten out of eleven transcripts were reduced to an expression level less than half of that of mock-injected embryos. Their average expression ranged from 7% to 34% of wildtype mRNA expression (Additional File 8). *Nv-CLANK-O* was not as effectively knocked down, but it was still reduced to ~64% of wildtype expression. Expression levels were monitored up to three days post eclosion, we found large variability in the behavior of the dsRNAs (Additional File 8). Some transcripts were immediately reduced, while others required a day to have an observable effect. Additionally, some transcripts were reduced for a number of days, while others quickly regained expression (Additional File 8).

We were encouraged that, despite the incomplete knockdown observed with qPCR, we still saw a significant increase in embryonic lethality. We sought to understand whether this lethality was due to disruptions in patterning in the early embryonic stages, where these CLANKs are expressed. We chose markers of dorsoventral patterning output that were strong, well understood, and represented the embryonic regions most sensitive to patterning disruption. Exactly how the observed disruptions come about is not known and will be the focus of future research.

Reduction of dorsolaterally expressed CLANK transcripts disrupts patterning specifically on the dorsal side

The *Nasonia* ortholog of *zerknüllt* (*Nv-zen*) is a well-established marker of the dorsal half of the embryo during normal development [13]. This zygotically expressed transcript is first observed in a broad stripe along the dorsal midline of the early blastoderm (Fig. 8A1, B1). This stripe stretches from the anterior to the posterior pole and is more or less equal in width and intensity throughout its domain. As the blastoderm continues to divide the domain narrows (Fig. 8A2, B2), and eventually retracts from the posterior pole as the blastoderm cellularizes and gastrulation begins (Fig. 8A3, B3). During gastrulation, *Nv-zen* marks the serosa until it begins to migrate and encompass the embryo (Fig. 8A4, B4). Because *Nv-zen* is the most consistently strongly expressed and most well characterized marker on the dorsal side of the embryo, it is an ideal marker to detect disruptions of patterning in this region of the embryo.

When individual laterally or dorsally expressed CLANK transcripts (*Nv-CLANK-A*, *-E*, *-F*, *-G*, *-H*, *-I*, *-K*, *-N*, *-O*) are knocked down, a number of changes in the expression of *Nv-zen* are observed (Fig. 8C1-F4, Additional File 9). First, in all transcripts tested (except *Nv-CLANK-K*) some embryos exhibited reduced levels of *Nv-zen* expression. When visible, the pattern of expression remained unchanged, however, the intensity of the stripe was much lower than control embryos that were processed in the same ISH experiment (Fig. 8D3-4). In other instances, the levels were too low to detect and the embryos appeared to be blank and absent of any *Nv-zen* expression (Fig. 8D1-2). For all of the knockdowns except for *Nv-CLANK-A* and *-K*, the increased frequency of reduced levels of *Nv-zen* expression was statistically significant ($p<0.05$) (Summarized in Fig. 9).

The second change observed is in the spatial domain of *Nv-zen*. The continuity of the *Nv-zen* expression domain is interrupted in a small proportion of embryos resulting from knockdown of *Nv-CLANK-E*, *-F*, *-N*, and *-O*. In mild cases, a small region adjacent to either the anterior or posterior pole is lacking expression (Fig. 8F1, F4), while the proximal pole and all distal regions appear unchanged. In more severe cases, larger regions, up to half the embryo (Fig. 8F2), or multiple regions throughout the embryo (Fig. 8F3) are lacking *Nv-zen* expression. This “incomplete stripe” phenotype was never observed in wildtype embryos. *Nv-CLANK-K* exhibited no abnormalities in *Nv-zen* expression (Summarized in Fig. 9).

To determine whether the spatial patterns of expression of the CLANKs are related to their regions of activity, the ventrally expressed *Nv-CLANK-M* was knocked down and, as would be expected if its function is restricted to its region of expression, the reduction of this ventrally expressed gene had no apparent effect on patterning the dorsal side of the embryo, as all observed embryos appeared phenotypically wildtype, displaying strong dorsal staining of *Nv-zen* (Summarized in Fig. 9).

Reduction of ventrolateral CLANK transcripts disrupts patterning, morphogenetic movements, and relative timing of embryonic events

Like *zen*, *twist* (*Nv-twi*) is a well-established marker of embryonic development in *Nasonia*, but for the ventral region of the embryo [13]. *Nv-twist* is first expressed in a thin stripe along the entire ventral midline of the early blastoderm (Fig. 10A1-B1). As the blastoderm undergoes additional divisions the stripe widens (Fig. 10A2-B2) before retracting at the anterior pole forming a pattern that resembles that of a slug (Fig. 10A3-

B3). This slug shape persists through cellularization and into the onset of gastrulation and marks the presumptive mesoderm specifically. The shape is lost once the mesoderm begins to internalize at the anterior end of the domain. This internalization progresses from anterior to posterior (Fig. 10B4) until the entire mesoderm is covered by neuroectoderm.

Ventral and laterally expressed CLANK transcripts (*Nv-CLANK-A*, *-E*, *-G*, *-H*, *-I*, *-K*, *-M*) were knocked down individually, and the expression pattern of *Nv-twi* was observed and characterized with *in situ* hybridization probes in a similar manner as with *Nv-zen* (Fig. 10C1-F4, Additional File 10). The reduction of these transcripts leads to a wide array of phenotypes.

The first group of phenotypes occurs in the early blastoderm when *Nv-twi* expands from a narrow to a wide ventral stripe (Fig. 10A1-B2). The reduction of *Nv-twi* expression is the first phenotype observed at this time point. Structurally these embryos appeared normal, and when present, the spatiotemporal domain of *Nv-twi* is unchanged. The frequency of embryos showing no or lower than normal *Nv-twi* expression (Fig. 10C1-9D2) appeared to be higher in all of the knockdowns except *Nv-CLANK-M*, was statistically significantly different from control for only for *Nv-CLANK-G and -I* (Fig. 11).

The second phenotype observed in the early blastoderm is a delay in the expansion of *Nv-twi* (Fig. 10F1). The expansion of the *Nv-twi* domain is stereotyped and occurs between nuclear cycle 10 and 11 [13] (compare Fig. 10A1, 10A2, 10E1). This delay phenotype is observed at a frequency higher than in wildtype embryos, after

knockdown of all of the lateral/ventral CLANK transcripts, but is only significantly higher for *Nv-CLANK-A*, *-E*, *-G*, *-M*. (Fig. 11).

Effects of CLANK knockdown become more frequent, severe, and varied in the late blastoderm stage, when *Nv-twi* is normally expressed in a ventral "slug" shaped domain (Fig. 10B3). Again, many embryos exhibit reduction in the expression levels of *Nv-twi*. Levels are sometimes reduced completely, as seen after knockdown of *Nv-CLANK-A*, *-E*, *-G*, *-I*, and *-K* (significantly increased frequency in all but *-A*, Fig. 12), or to a much lower level than observed in wildtype embryos (*Nv-CLANK-G*, Fig. 10F2, significantly increased frequency, Fig. 12).

Nv-CLANK-A, *-E*, *-G*, *-H*, *-I*, or *-M* knock downs all lead to a disruption in the slug shaped domain of *Nv-twi*. Normally this pattern has very sharp, straight lateral borders and a very distinct forking at its anterior end. The sharpness and straightness of the lateral borders are affected at a low, but consistent, frequency after *Nv-CLANK-A*, *-E*, *-G*, *-H*, *-I*, and *-M* knockdown (Fig. 10D3). In rarer cases, *Nv-twi* domains where the anterior fork did not resolve are observed (only in *Nv-CLANK-G* and *-M*, Fig. 10F3). Finally, in some embryos the edges of the *Nv-twi* domain remain unchanged, but within the domain, large patches of cells lack *Nv-twi* expression (Fig. 10D4). The size and number of these patches vary from embryo to embryo within and between knock down conditions.

This “patchy” phenotype was observed after knockdown of all seven ventral/lateral *CLANKs*, and at a much lower frequency in control embryos. However the difference in frequency of this phenotype was significantly only in *Nv-CLANK-M*

knockdowns relative to control (Fig. 12). The “messy” border and missing anterior fork phenotypes were never observed in wildtype embryos (Fig. 12).

The last time point where we looked for disruption is during gastrulation. Again, embryos were observed that lack positive staining for *Nv-twi* expression as with the two early stages of development (data not shown); however, this also occurred in wildtype embryos, and only the knockdown of *Nv-CLANK-K* resulted in a frequency of this phenotype significantly higher than what is expected (Fig. 13).

More interestingly, some knockdowns disrupted the morphogenetic movements of gastrulation. Normally mesoderm internalization proceeds from anterior to posterior in *Nasonia* [13] (Fig. 10B4). In rare instances, the mesoderm was observed internalizing posteriorly to anteriorly (Fig. 10F4) or in a random, disorganized manner (Additional file 10D, Q, Y, Z, EE) when *Nv-CLANK-A*, *-G*, *-K*, or *-M* are knocked down. While never observed in control embryos (or in the myriad normal embryos observed in other experiments), this phenotype occurred at the lowest frequency of all that have been described, and in no condition is the frequency statistically significant (Fig. 13).

To again test whether the spatial expression of the CLANKs is correlated with the location of their phenotypic effects, knockdown embryos from *CLANKs* expressed on the dorsal half of the embryo (*Nv-CLANK-F*, *-N*, *-O*) were also examined for changes in *Nv-twi* expression. As expected, the loss (or reduction) of these dorsal transcripts had no effect on patterning of the ventral side of the embryo. All observed embryos appeared phenotypically wildtype, displaying strong *Nv-twi* staining (*Nv-CLANK-F*: Fig. 11-13 *Nv-CLANK-N, -O*: data not shown).

521 In summary, all of the 11 tested genes showed an increase in embryonic lethality
522 compared to control. We could then show that markers of DV cell fates are disrupted
523 spatially (for both *Nv-twi* and *Nv-zen*) and temporally (*Nv-twi* expansion) by knockdown
524 of different CLANKs. This shows that these novel components of the DV GRN are
525 functionally integrated and are important in producing a stable and reproducible
526 patterning output.
527 The above results led us to wonder how long these genes have been a part of DV
528 patterning in the wasp lineage to *Nasonia.* Are all of these genes unique recent
529 additions to *Nasonia* DV GRN, or do some of them have a longer history in the wasp
530 lineage?
531 **Discovery of CLANKs in the wasp *Melittobia digitata***
532 The second approach to understand the developmental and evolutionary
533 significance of this gain of DV expression in *Nasonia* was to examine the function and
534 expression of CLANKs in other species. This will help to understand how these genes
535 have been functionally integrated into developmental processes.
536 As we described above, it appears that CLANKs are an ancestral and unique
537 feature of the Superfamily Chalcidoidea. We have chosen to develop *Melittobia digitata*,
538 a representative of the Family Eulophidae (separated from *Nasonia* by about 90 million
539 years of independent evolution [28]), as a comparative model. *Melittobia* is attractive
540 because it is easily reared in the lab on the same hosts as *Nasonia*, its mode of
541 embryogenesis is rather similar to *Nasonia,* allowing for more straightforward
542 comparisons of expression patterns, and it adds an important phylogenetic sampling

point in understanding the evolution of development within the megadiverse Chalcidoidea.

We sequenced and assembled an embryonic transcriptome from *Melittobia*, and then searched for potential orthologs of *Nasonia* CLANKs within this transcriptome using local BLAST [29]. The sequences of the *potential Melittobia* CLANK homologs are presented in Additional File 5, and were used to generate antisense probes to assess expression patterns (Fig. 14) and in phylogenetic analysis to assess their relationships among themselves and with *Nasonia* CLANKs (Fig. 15).

Characterization of *Melittobia* CLANK expression

Since there was only weak evidence for direct orthology of *Melittobia* sequences to the *Nasonia* DV CLANKs, we considered all of the *Melittobia* genes we found to be potential homologs and assessed their expression. Ten of the seventeen Md-CLANKs (*Md-CLANK-A*, *-B*, *-D*, *-H*, *-I*, *-I2*, *-J*, *-K*, *-L*, *-N*) we identified as potential homologs of the DV Nv-CLANKS were not expressed differentially along the DV axis (Additional File 11).

In contrast, *Md-CLANK-C* again has dynamic expression in early and gastrulating embryos (Fig. 14A1-A6). It is absent in pre-blastoderm embryos (Fig. 14A1), then is initially expressed in three bands along the AP axis of the embryo (Fig. 14A2). The strongest and most complete is near the posterior pole. Expression increases in strength and size forming a lateral domain that almost encapsulates the whole embryo. Expression is lacking at the two poles, along the dorsal midline, and along most of the ventral midline (Fig. 14A3-A3'). This lateral domain then retracts into two discrete bands of expression (Fig. 14A4-A4'). The anterior most bands then disappears, leaving just

566 one posterior band (Fig. 14A5). Throughout this retraction the lack of staining at the
567 dorsal midline and poles persists, however, there is staining at the ventral midline in the
568 two bands (compare Fig. 14A3-A4'). During gastrulation this posterior band of
569 expression slowly fades until expression is lacking throughout the embryo (Fig. 14A6).
570 This pattern is quite similar to previous patterns seen in *Nv-CLANK-G* and *-H*. These
571 genes are in a phylogenetic cluster of *Nasonia* CLANKs that is sister to a cluster of
572 exclusively *Melittobia* CLANKs that contains *Md-CLANK-C* (Fig. 15).
573 *Md-CLANK-E* and *-E2* also have low levels of ubiquitous expression in early
574 embryos and no expression during gastrulation (Fig. 14B1, B4, C1, C4). However, in
575 blastoderm embryos, both are expressed in a stripe along the dorsal midline. The stripe
576 is dynamic in expression levels and size along the entire AP axis for both *CLANKs*.
577 Expression appears to originate at the anterior pole and fill in in a discontinuous manner
578 until the entire dorsal midline is exhibits expression (Fig. 14B2, B3, C2, C3).
579 Satisfyingly, these two genes cluster phylogenetically with *Nv-CLANK-O* (Fig. 15), which
580 is expressed in an almost identical narrow dorsal stripe (Fig. 5B1-B2). This strongly
581 indicates that the common ancestor of these genes was expressed dorsally and that
582 this pattern has persisted for 90 million years.
583 Expression of *Md-CLANK-F* and *-F2* is dynamic during blastoderm stages of
584 development. Early pre-blastoderm and blastoderm-staged embryos have light
585 ubiquitous expression (Fig. 14D1, E1). Expression is then increased in the yolk of the
586 syncytial blastoderm (Fig. 14D2, E2), reduced to low levels throughout, and then
587 localized in a small patch of the dorso-posterior of the embryo (Fig. 14D3,E3) before
588 quickly being lost again in cellularized blastoderm stage and gastrulating embryos (Fig.

589 14D4,E4). This pattern has no clear counterpart in the *Nasonia* genes we have
590 examined.

591 In early blastoderm embryos, *Md-CLANK-G* expression appears to be weak and
592 ubiquitous, with slightly higher expression on the ventral half of the embryo (Fig. 14F1).
593 Expression increases in intensity forming a stripe along the ventral midline, widest in the
594 anterior third of the embryo and narrowing in the posterior third (Fig. 14F2-F3). At the
595 onset of gastrulation, expression is strongest and resembles the characteristic slug
596 shaped domain of *Nv-twist,* before staining is lost completely (Fig. 14F4). This pattern
597 does not develop similarly to the ventrally expressed *Nv-CLANK-M* (Fig. 6A1-A3), and
598 we do not consider it homologous.

599 The phylogenetic analysis revealed that there has likely been large-scale
600 duplication and divergence (and/or gene conversion) in both wasp lineages (Fig. 15).
601 Most of the *Nasonia* DV CLANKs cluster in two distinct clades on either side of the
602 basal split of this protein tree. Similarly, most of the *Melittobia* proteins cluster together,
603 or with the "off-target" *Nasonia* CLANKs that are not involved in DV patterning (Fig. 15).
604 There are only a handful of cases that indicate clear orthology between a *Nasonia* DV
605 CLANK and a *Melittobia* CLANK. *Nv*-CLANK-O clusters strongly with *Md*-CLANK-E1
606 *and* -E2*,* while *Nv*-CLANK-K clusters with *Md*-CLANK-M and -D. For the others, more
607 complex evolutionary histories, involving ancestral genes that duplicated and diverged
608 multiple times in both lineages after their separation must be considered. Thus, we
609 propose that *Nv*-CLANK-G, -A, -H, -L, and -C derived the same common ancestral gene
610 as *Md*-CLANK-B, -C, -G, -I1, -I2, and -K. On the other hand, it is unclear how to relate
611 *Nv*-CLANK-F, -N, -M, -D, -I, and -J to *Melittobia* counterparts.

612 Horizontally transferred and duplicating genes have been shown to be subject to
613 complex processes of molecular evolution that make defining ancestry of genes
614 particularly difficult [30], which might explain the difficulty in defining orthology in these
615 relatively closely related species. In addition, potentially missing orthologs may have
616 been misassembled in our transcriptome, and thus not picked up by BLAST analysis.
617 However, the genes we now have in hand are already quite informative as to the
618 functional evolution of this gene family.
619

620 **Discussion**

621 In this paper we have shown that a group of genes that came about from multiple
622 rounds of gene duplication and divergence events, potentially following one or more
623 HGT events, are stably and functionally integrated into the embryonic DV patterning
624 GRN of the wasp *Nasonia.* Furthermore, we provide evidence that some of the
625 functionally integrated genes have been participating in developmental processes for a
626 long period, extending back at least 90 million years to the common ancestor of *M.*
627 *digitata* and *N. vitripennis*. These results raise myriad questions about the origin and
628 fate of horizontally transferred genes, why they are sometimes maintained, and how
629 GRNs change in the course of incorporating these invading genes.

630 **Incorporation, duplication, and diversification of DV CLANKs**

631 Ankyrin-repeat motifs (ANK) are important for protein-protein interactions and are
632 commonly found in proteins across many species [23, 31]. Sequencing of the *Nasonia*
633 genome uncovered that it contains the largest number of genes coding for ANK proteins
634 of any insect [17]. Among this large number of ANK domain containing are orthologs to

635 genes that are conserved features of insect genomes. However, the vast majority
636 (~170) are orphan genes without clear orthologs in other insects that we have termed
637 CLANKs in this manuscript. A clue to the microbial origin of the CLANKs was the
638 discovery of PRANC domains at the C-termini of some of the proteins. The PRANC
639 domain is found in *Wolbachia*, its bacteriophage, poxviruses and various other bacteria,
640 and its presence strongly indicated HGT from *Wolbachia* into the genome of an
641 ancestor of *Nasonia.* Our results using PSI-BLAST to identify cryptic similarity of the C-
642 termini of CLANKS that do not have annotated PRANC domains to C-termini of
643 *Wolbachia* PRANC and ankyrin domain containing proteins further bolsters this case.
644 Our observations indicate that the large number of CLANKs in *Nasonia* is the
645 result mostly of duplication and divergence of genes present in the most recent common
646 ancestor of the Chalcidoidea, rather than repeated HGTs within the *Nasonia* lineage
647 after the split among the families. This is based on our observation that the proteins are
648 highly diverged from each other and from any presumed ancestor found in *Wolbachia*.
649 In addition, the presence of introns in almost all of the sequences, their clear integration
650 into the transcriptional regulation milieu of the *Nasonia*, and their dispersal throughout
651 the genome strongly indicate that at least the 15 genes we detected as DV regulated
652 have a long history in wasp genomes, and likely have arisen from duplication and
653 divergence processes.
654 Apparently more recent HGT events may give clues about the origin and
655 evolution of the CLANKs. For example, in the ant *Pseudomyrmex gracilis* [32], and the
656 bee *Ceratina* [33] we find dozens of Ankyrin domain containing genes that cluster
657 together in phylogenetic analyses, and also cluster tightly with Wolbachia in these same

658 analyses (Additional figures 3 and 4). We cannot exclude that multiple (or no) HGT
659 events occurred in the lineage leading to *Nasonia* to give the full complement of CLANK
660 genes in this wasp, and analyses to determine such facts should be an area of
661 considerable effort in the future.

662 **Why were the ancestors of the *Nasonia* DV CLANKs maintained?**

663 The chances of a gene horizontally transferred from a prokaryote to a eukaryote
664 to be maintained in a genome is likely to be very low, since not only must it gain the
665 ability to be activated and processed by the eukaryotic transcriptional machinery, but it
666 should also quickly gain a function in its new milieu. If these conditions are not met
667 quickly, random mutations will accumulate and without selection, will eventually destroy
668 protein-coding capacity of the transferred sequence (this is true whether a new gene
669 arises by duplication, HGT, or de novo [34-38].

670 Since ankyrin domains are protein binding domains [31, 39, 40] and direct
671 protein-protein interaction is thought to be an important pre-cursor for proteins to gain
672 new function [41, 42], ankyrin domain encoding genes like the CLANKs may be
673 predisposed to gain function in new environments. Furthermore, assuming the ancestral
674 CLANK possessed a PRANC domain, it could have a ready-made interaction partner in
675 the form of the Nf-κB homolog Dorsal, which has conserved roles in innate immunity
676 and embryonic patterning throughout insects [43].

677 **Why were CLANKS integrated into DV patterning?**

678 One of the most surprising results of our previous analysis of DV patterning
679 genes in *Nasonia* was the discovery of so many CLANKs with distinct and unique
680 expression patterns. Their potential as important regulators of the Toll/Dorsal pathway

was quite exciting, especially as there are still major open questions about how the Toll/Dorsal pathway interacts with BMP signaling to pattern the *Nasonia* embryo [14]. In Poxviruses, PRANC domain containing genes are known to inhibit the activation of the NF-kB pathway, hijacking the innate immune system within their hosts [44]. Additionally, the PRANC domain has been described as being very similar to F-box domains [45], a domain that is known to induce ubiquitination of IkB, its degradation, and the activation of NFkB [46]. These associations with NFkB/IkB are very interesting because while these proteins function as innate immune responses in most mammals, they have been co-opted to have a function in DV patterning of the embryo of higher insects (Dorsal/Cactus), including *Nasonia*. Therefore, it is possible that these CLANKs were incorporated into the DV pathway because they already had previously established interaction domains with proteins within the pathway (Dorsal/Cactus).

That being said, we have made preliminary observations that indicate that integration of CLANKs is pervasive throughout *Nasonia* development. First, our result reported here, that some CLANK pRNAi knockdowns lead to maternal lethality indicates that some of the DV CLANKs have additional roles in adult organisms. In addition, many of our DV CLANKs are clearly also regulated along the anterior-posterior axis, which might indicate that additional CLANKs might also play important roles in the zygotic GRN patterning this axis. Finally, we have identified additional CLANKs showing maternal mRNA localization to both the anterior and posterior poles of the oocyte/early embryo, indicating roles in establishing AP polarity and specifying germ cells, and CLANKs that are specifically upregulated in either male or female embryos, indicating a role in sex determination (JAL, personal observations). Further functional

704 approaches will be undertaken to assess functional integration of CLANKs into these,
705 and additional GRNs.

706 **What are the molecular roles of DV CLANKs?**

707 While we propose that interaction with Toll/Dorsal signaling may have been
708 important in the initial integration and stabilization of the CLANKS into the ancestral
709 Chalcidoidean genome, , it is not clear that this interaction has been maintained for the
710 modern *Nasonia* CLANKs. The "incomplete stripe" of *Nv-zen* seen after knockdown of
711 *Nv-CLANK-E*, *-F*, *-N* and *-O* is reminiscent of weak knockdowns of BMP components in
712 *Nasonia* ([14], JAL personal observation), whereas knockdown of Toll has no effect on
713 *Nv-zen*. In addition, the strong reduction of intensity of *Nv-zen* could be a disruption of
714 the BMP signaling, or transcription of target genes downstream of this pathway.

715 The results where the expansion of the *Nv*-twist domain is delayed could be
716 ascribed to a disruption of Toll signaling. However, our previous work indicated Toll
717 signaling specifies the initial narrow stripe, while the expansion is mediated by zygotic
718 factors. Similarly, the later disrupted border is likely caused by disruption of interactions
719 of zygotic DV patterning genes, rather than Dorsal itself.

720 At some level, it is not surprising that the CLANKs we find today have not
721 maintained their hypothetical ancestral interaction partners, given the strong divergence
722 of these genes from each other at the amino acid level, and their 150 million years of
723 evolution. Understanding the molecular interactions that mediate the function of the
724 CLANKs will be a high priority in the coming years.

725

726 **When were developmental roles for CLANKs fixed in Chalcidoidea, and how**
727 **common is recruitment of these genes for developmental processes?**

728 Our curiosity about the evolutionary history of the functional integration of
729 CLANKs into the *Nasonia* DV patterning network led us to establish the wasp *Melittobia*
730 as a satellite model organism. Our results indicate on one hand that there is evidence
731 that CLANKs have been integrated into the DV patterning GRN of some wasps for at
732 least 90 million years (since the divergence of the Pteromalid and Eulophid Families)
733 [28], based on the similarity of expression of two sets of CLANK homologs. On the other
734 hand, we also revealed lineage specific expression patterns for genes in both species,
735 indicating that functional integration of CLANKs is an ongoing process. Sampling wasps
736 in other Families in the Chalcidoidea will be necessary to pinpoint the origin of functional
737 integration of DV CLANKs, and to provide directionality to the changes (i.e., are some
738 functional CLANKs being lost in some lineages, or is the pattern more due to
739 independent gains?). It is likely that an unbiased approach to identify DV regulated
740 genes in *Melittobia* similar to the one we took in *Nasonia* [15], and broad application of
741 this approach in a defined phylogenetic entity would allow for a high resolution
742 characterization of the evolutionary history of this gene family.

743 **Conclusion**

744 Our knockdowns of DV CLANKs showed that these genes significantly
745 participate in ensuring successful embryogenesis and the formation of viable larvae.
746 Further we showed that a potential role of these genes is to ensure proper
747 establishment of cell fates along the DV axis. However, in neither the patterning nor the
748 viability roles do the CLANKs appear to be absolutely essential. Rather, we propose

749 that CLANKs act to constrain fluctuations in early development, and that their loss can
750 lead to variable fluctuations in patterning which in only rare cases lead to lethality. The
751 source of the fluctuations could be environmental, genetic, or a combination of the two
752 [6, 47, 48]. Our results show that even a very modest contribution to stability of
753 development may lead novel GRN components to be maintained over significant
754 evolutionary time periods. Thus, our results can be extrapolated to any potential novel
755 components of GRNs, whether they originate from HGT, de novo genes, and co-option
756 of existing genes into a new network.

757 **Methods**

758 **BLASTs and sequence characterization**

759 Our starting sequences were from the *Nasonia* annotation 2.0 [49]. Their NCBI
760 counterparts were found by BLAST against the nt database [50]. These results provided
761 the corresponding NCBI Reference Sequence Accession information for the *Nasonia*
762 2.0 annotations, which was used for downstream analyses of protein function.
763 Other similarity searches used blastp and searched the non-redundant protein
764 database and default parameters, except cases where searches were limited to a single
765 species.
766

767 **Protein sequence alignment, phylogeny, and conserved domain analysis**

768 Protein sequences corresponding to these fifteen transcripts were then submitted
769 to search for conservation at the amino acid level and to rendered a phylogenetic tree
770 using “One Click Robust Phylogenetic Analysis” (http://phylogeny.lirmm.fr/) [25]. “One
771 Click” parameters were as followed: Data & Settings (Gblocks not used), MUSCLE
772 Alignment (-SeqType Protein), PhyML Phylogeny (Substitution model: WAG), TreeDyn
773 Tree Rendering (Reroot using mid-point rooting, Branch annotation: Branch support
774 values). Finally, in order to gain family, domain and repeat information for each
775 transcript, we analyzed each protein using the Interpro Protein sequence analysis &
776 classification software (InterProScan5) (https://www.ebi.ac.uk/interpro/) [26].

Phylogenetic analysis of the top 100 BLAST hits to each CLANK was performed at http://www.trex.uqam.ca/ [51]. Alignments were performed using MUSCLE [52], alignments were manipulated for correct file type using AliView [53], relationships were inferred with RAxML [54] and trees were drawn and edited in FigTree: (http://tree.bio.ed.ac.uk/software/figtree/). Default parameters were used.

PSI_BLAST was performed using sequence downstream of annotated ankyrin domains in *Nasonia* CLANKs. 2 iterations were performed for each gene, except CLANK-L, which required 4 iterations to find sequences outside of *Nasonia* and *Trichomalopsis*. Only sequences above threshold were used to seed the next iteration. All aligning sequences were downloaded as FASTA files. These were aligned all to eachother using MUSCLE implemented at T-rex as above, and the resulting alignments were used for phylogenetic analysis as described above.

RNA interference, screening, and embryo collection

Yellow AsymCx (wild-type, cured of *Wolbachia*) pupa were injected with dsRNA (~1µg/mL in water) designed against each of the transcripts with significant DV expression as described in [27]. Injected pupae were allowed to eclose and lay eggs. Overnight egg lays were collected and plated onto 1% PBS agar plates. Embryos were aged 24 hours at 28^{o}C and then screened for embryonic lethality. Mock, water-injected embryos were also collected, plated, and screened as a control. Nearly all water-injected embryos are predicted to hatch within 24 hours and develop into crawling, feeding, larva. A small number of embryos will fail to hatch, and instead development will arrest at the embryonic stage. We define this failure to hatch as embryonic lethality. If the transcript we knockdown is predicted to be vital for embryonic development, we predict that a higher percent of injected embryos will exhibit this embryonic lethal phenotype compared to mock-injected embryos.

Average embryonic lethality was calculated and plotted as a bar graph for each condition. Standard deviation was used to calculate standard error, and T-tests ($p < 0.05$) were used to test for significance. Box plots were also created to provide visualization of the distribution of observed lethality within each conditional population.

Timed egg lays were also conducted to collect 3-7 hours (at 28^{o}C) embryos from

each knockdown condition. This time span corresponds to the developmental stages we know these transcripts are differentially expressed (the penultimate syncytial division through the beginning of gastrulation) Embryos were fixed and then processed for in situ hybridization or qPCR.

Characterization of RNA localization (*in situ* hybridization)

in situ hybridization was performed using standard protocols [13, 15] on 0-24 hour, wildtype, AsymCX embryos in order to characterize normal expression patterns of each CLANK transcript during embryogenesis (specific details available upon request). Embryos were imaged at 20X magnification on Zeiss widefield, compound epi-fluorescent microscope.

For knockdown experiments, *in situ* protocols were repeated on the 3-7 hours (28^{o}C) knockdown embryos and on 3-7 hour mock-injected embryos. Anti-sense probes were generated from primers specific to *Nv-twi* and *Nv-zen*. Knockdown phenotypes were described based on their divergence from mock-injected expression of *Nv-twi* and *Nv-zen*, and their frequency of occurrence was calculated and compared to mock-injected phenotype frequencies. Raw frequency counts were converted to percentages (out of 100) and Fisher's Exact Test was used to determine if a given phenotypic frequency observed in knockdown embryos was significantly different from the frequency of that phenotype in mock-injected embryos ($p < 0.05$).

Qualitative Polymerase Chain Reaction (qPCR)

RNA was isolated from 3-7 hours (28^{o}C) embryos using standard TRIzol-based protocols (Ambion 15596018) and converted into cDNA using the Protoscript First Strand cDNA synthesis kit (NEB 63001), controlling for total RNA input. Two cDNA replicates were synthesized per condition. cDNA was synthesized in this manner for each condition for three consecutive days post eclosure of the injected wasps.

To assess knockdown, we performed qPCR on knockdown embryos in parallel with mock-treated embryos. These were carried out using primers specific to the transcript of interest while using the housekeeping gene, *rp49*, as a control. 20µL per well PCR reactions were assembled using the PowerUp SYBR Green Master Mix

(Applied Biosystems: A25742)(2X MM, 800nM of each primer, cDNA, RFH2O). We performed the reactions in triplicate using the following parameters: (50^{o}C for 2', 95^{o}C for 2', 40 cycles of (95^{o}C for 15 sec, 60^{o}C for 60 sec, plate read, 72^{o}C for 60 sec, plate read), 95^{o}C for 2', gradient 60^{o}C→95^{o}C (0.2^{o}C for 1 sec).

Average C_T was calculated by combining triplicates from both cDNA replicates for each condition. Knockdown average C_T 's were then normalized to mock injected RP49 levels. Knockdown Delta C_T 's were calculated and expressed as a relative expression (percentage of wildtype expression). Relative expression was calculated for each condition per day (up to three days) and an average relative expression was calculated over the three-day span. Standard deviation was used to calculate standard error, and T-test ($p < 0.05$) were used to test for significance.

Melittobia sequences

Total mRNA (1 ug) libraries were created from various developmental time points (ovaries, pre-, early-, late-blastoderm embryos, male-, female-yellow pupa) of the wasp *Melittobia digitata* using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB #E7420) in conjunction with NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490). Libraries were validated and quantified before being pooled and sequenced on an Illumina HiSeq 2000 sequencer with a 100 bp paired-end protocol. Sequences were de novo assembled using Trinity on a Galaxy Portal. (Currently in preparation for publication, specific details available upon request). *Melittobia* RNAseq data is available in the BioSample database under accession numbers [SAMN08361226, SAMN08361227, SAMN08361228].

Melittobia ortholog discovery

We used a de novo assembled, unannotated embryonic *Melittobia digitata* Transcriptome as a database for local tBLASTn (*Nv*-protein → *Md*-mRNA) within the Geneious program (https://www.geneious.com/) [29] to search for orthologs to the

867 *Nasonia CLANK* genes. A reciprocal BLAST was done on the top hits. If the reciprocal
868 BLAST resulted in the *Nasonia* sequence that was input into the query being returned
869 as the top hit, the hit was considered a strong ortholog candidate. If it did not correctly
870 BLAST back to the input sequence, the off-target *Nasonia* protein sequence returned
871 was collected to be input in to downstream phylogenetic analysis. *Melitobia* transcript
872 sequences were then translated into amino acid sequences via an online Translate tool
873 (ExPASy, https://web.expasy.org/translate/). Longest ORFs sequences were confirmed
874 to align with tBLASTn hit sequences, and then were collected for phylogenetic analysis.
875 Input *Nasonia*, off-target *Nasonia*, and all potential *Melittobia* ortholog protein
876 sequences were then submitted to "One Click Robust Phylogenetic Analysis"
877 (http://phylogeny.lirmm.fr/) [25] and trees were rendered relating each chalcid species
878 orthologs to the *Nasonia* sequences as described earlier in our methods.
879
880 ***Melittobia in situ* hybridization**
881 Embryos collection, processing, and *in situ* protocol were developed and
882 performed in a manner similar to *Nasonia* protocols with minor modifications. (Currently
883 in preparation for publication, specific details available upon request).
884
885 **List of abbreviations**
886 **ANK**: Ankyrin-repeat motifs
887 **CLANK**s: **C**halcidoidea **L**ineage specific **ANK**yrin domain encoding genes
888 **DV**: dorsoventral
889 **GRN**: Gene regulatory networks
890 **HGT**: horizontal gene transfer
891 ***htl*:** heartless
892 **ISH**: *in situ* hybridization
893 **PRANC**: **P**ox proteins **R**epeats of **AN**kyrin, **C**-terminal domain
894 **pRNAi**: parental RNA interference

qPCR: Quantitative PCR

sna*: snail*

twi*: twist*

zen*: zerknüllt*

zfh**:** *zinc finger homeodomain*

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Melittobia RNAseq data is available in the BioSample database under accession numbers [SAMN08361226, SAMN08361227, SAMN08361228], the SRA database under accession number SRP129036, and the BioProject database under accession number PRJNA429828.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors were supported under NIH Grant # 1R03HD087476, 1R03HD078578 (JAL) and startup fund from University of Illinois at Chicago.

Authors' contributions

DP provided assistance in experimental design and performed all the experiments. JAL designed the experiments and supervised students. All authors contributed to writing the manuscript and approved the final manuscript.

Acknowledgements

We would like to like to acknowledge the DNA Services Facility at UIC for assistance in transcriptome sequencing. We would like to thank Sarah Bordenstein and an anonymous reviewer for helpful suggestions and discussion on this manuscript.

REFERENCES

1. Levine, M. and E.H. Davidson, *Gene regulatory networks for development.* Proc Natl Acad Sci U S A, 2005. **102**(14): p. 4936-42.
2. Stathopoulos, A. and M. Levine, *Genomic regulatory networks and animal development.* Dev Cell, 2005. **9**(4): p. 449-62.
3. Moczek, A.P., *Developmental capacitance, genetic accommodation, and adaptive evolution.* Evol Dev, 2007. **9**(3): p. 299-305.
4. Wilkins, A.S., *Canalization: a molecular genetic perspective.* Bioessays, 1997. **19**(3): p. 257-62.
5. Nijhout, H.F., *Control mechanisms of polyphenic development in insects - In polyphenic development, environmental factors alter same aspects of development in an orderly and predictable way.* Bioscience, 1999. **49**(3): p. 181-192.
6. von Dassow, G., et al., *The segment polarity network is a robust developmental module.* Nature, 2000. **406**(6792): p. 188-92.
7. Moczek, A.P., *On the origins of novelty in development and evolution.* Bioessays, 2008. **30**(5): p. 432-47.
8. Sommer, R.J., et al., *The Pristionchus HOX gene Ppa-lin-39 inhibits programmed cell death to specify the vulva equivalence group and is not required during vulval induction.* Development, 1998. **125**(19): p. 3865-3873.
9. Abouheif, E. and G.A. Wray, *Evolution of the gene network underlying wing polyphenism in ants.* Science, 2002. **297**(5579): p. 249-252.
10. Wotton, K.R., et al., *Quantitative System Drift Compensates for Altered Maternal Inputs to the Gap Gene Network of the Scuttle Fly Megaselia abdita.* Elife, 2015. **4**.
11. Wiegmann, B.M., et al., *Single-copy nuclear genes resolve the phylogeny of the holometabolous insects.* BMC Biol, 2009. **7**: p. 34.
12. Davis, G.K. and N.H. Patel, *Short, long, and beyond: molecular and embryological approaches to insect segmentation.* Annu Rev Entomol, 2002. **47**: p. 669-99.

13. Buchta, T., et al., *Patterning the dorsal-ventral axis of the wasp Nasonia vitripennis.* Dev Biol, 2013. **381**(1): p. 189-202.
14. Ozuak, O., et al., *Dorsoventral Polarity of the Nasonia Embryo Primarily Relies on a BMP Gradient Formed without Input from Toll.* Current Biology, 2014. **24**(20): p. 2393-2398.
15. Pers, D., et al., *Global analysis of dorsoventral patterning in the wasp Nasonia reveals extensive incorporation of novelty in a regulatory network.* Bmc Biology, 2016. **14**.
16. Bordenstein, S.R. and S.R. Bordenstein, *Eukaryotic association module in phage WO genomes from Wolbachia.* Nature Communications, 2016. **7**.
17. Werren, J.H., et al., *Functional and evolutionary insights from the genomes of three parasitoid Nasonia species.* Science, 2010. **327**(5963): p. 343-8.
18. Xiao, J.H., et al., *Obligate mutualism within a host drives the extreme specialization of a fig wasp genome.* Genome Biology, 2013. **14**(12).
19. Lindsey, A.R.I., et al., *Comparative genomics of the miniature wasp and pest control agent Trichogramma pretiosum.* Bmc Biology, 2018. **16**.
20. Penz, T., M. Horn, and S. Schmitz-Esser, *The genome of the amoeba symbiont "Candidatus Amoebophilus asiaticus" reveals common mechanisms for host cell interaction among amoeba-associated bacteria.* Virulence, 2010. **1**(6): p. 541-545.
21. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.
22. Machado, C.A., et al., *Phylogenetic relationships, historical biogeography and character evolution of fig-pollinating wasps.* Proc Biol Sci, 2001. **268**(1468): p. 685-94.
23. Jernigan, K.K. and S.R. Bordenstein, *Ankyrin domains across the Tree of Life.* Peerj, 2014. **2**.
24. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Molecular Systems Biology, 2011. **7**.
25. Dereeper, A., et al., *Phylogeny.fr: robust phylogenetic analysis for the non-specialist.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W465-9.
26. Jones, P., et al., *InterProScan 5: genome-scale protein function classification.* Bioinformatics, 2014. **30**(9): p. 1236-1240.
27. Lynch, J.A. and C. Desplan, *A method for parental RNA interference in the wasp Nasonia vitripennis.* Nat Protoc, 2006. **1**(1): p. 486-94.
28. Peters, R.S., et al., *Evolutionary History of the Hymenoptera.* Current Biology, 2017. **27**(7): p. 1013-1018.
29. Kearse, M., et al., *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.* Bioinformatics, 2012. **28**(12): p. 1647-9.
30. Hao, W. and J.D. Palmer, *HGT turbulence: Confounding phylogenetic influence of duplicative horizontal transfer and differential gene conversion.* Mob Genet Elements, 2011. **1**(4): p. 256-261.
31. Mosavi, L.K., et al., *The ankyrin repeat as molecular architecture for protein recognition.* Protein Science, 2004. **13**(6): p. 1435-1448.

32. Rubin, B.E.R. and C.S. Moreau, *Comparative genomics reveals convergent rates of evolution in ant-plant mutualisms.* Nature Communications, 2016. **7**.
33. Rehan, S.M., et al., *The Genome and Methylome of a Subsocial Small Carpenter Bee, Ceratina calcarata.* Genome Biol Evol, 2016. **8**(5): p. 1401-10.
34. Chen, S., et al., *Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene.* EMBO J, 2012. **31**(12): p. 2798-809.
35. Zhang, W., et al., *New genes drive the evolution of gene interaction networks in the human and mouse genomes.* Genome Biol, 2015. **16**: p. 202.
36. Lynch, M., et al., *The probability of preservation of a newly arisen gene duplicate.* Genetics, 2001. **159**(4): p. 1789-804.
37. Lynch, M. and J.S. Conery, *The evolutionary fate and consequences of duplicate genes.* Science, 2000. **290**(5494): p. 1151-5.
38. Assis, R. and D. Bachtrog, *Neofunctionalization of young duplicate genes in Drosophila.* Proc Natl Acad Sci U S A, 2013. **110**(43): p. 17409-14.
39. Sedgwick, S.G. and S.J. Smerdon, *The ankyrin repeat: a diversity of interactions on a common structural framework.* Trends in Biochemical Sciences, 1999. **24**(8): p. 311-316.
40. Bork, P., *Hundreds of Ankyrin-Like Repeats in Functionally Diverse Proteins - Mobile Modules That Cross Phyla Horizontally.* Proteins-Structure Function and Genetics, 1993. **17**(4): p. 363-374.
41. Lynch, M., *The Evolution of Multimeric Protein Assemblages.* Molecular Biology and Evolution, 2012. **29**(5): p. 1353-1366.
42. Capra, J.A., K.S. Pollard, and M. Singh, *Novel genes exhibit distinct patterns of function acquisition and network integration.* Genome Biology, 2010. **11**(12).
43. Moussian, B. and S. Roth, *Dorsoventral axis formation in the Drosophila embryo - Shaping and transducing a morphogen gradient.* Current Biology, 2005. **15**(21): p. R887-R899.
44. Chang SJ, H.J., Sonnberg S, Chiang CT, Yang MH, Tzou DL, Mercer AA, Chang W., *Poxvirus host range protein CP77 contains an F-box-like domain that is necessary to suppress NF-kappaB activation by tumor necrosis factor alpha but is independent of its host range function.* J Virol., 2009. **83**: p. 4140–52.
45. Mercer, A.A., S.B. Fleming, and N. Ueda, *F-box-like domains are present in most poxvirus ankyrin repeat proteins.* Virus Genes, 2005. **31**(2): p. 127-133.
46. Spencer E., J.J., Chen Z.J., *Signal-induced ubiquitination of IκBα by the F-box protein Slimb/β-TrCP.* Genes Dev., 1999. **13**: p. 284–294.
47. Kitano, H., *Biological robustness.* Nat Rev Genet, 2004. **5**(11): p. 826-37.
48. Gavin-Smyth, J., et al., *A genetic network conferring canalization to a bistable patterning system in Drosophila.* Curr Biol, 2013. **23**(22): p. 2296-302.
49. Rago, A., et al., *OGS2: genome re-annotation of the jewel wasp Nasonia vitripennis.* Bmc Genomics, 2016. **17**.
50. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
51. Boc, A., A.B. Diallo, and V. Makarenkov, *T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W573-9.

52. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
53. Larsson, A., *AliView: a fast and lightweight alignment viewer and editor for large datasets.* Bioinformatics, 2014. **30**(22): p. 3276-3278.
54. Stamatakis, A., *RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.* Bioinformatics, 2006. **22**(21): p. 2688-90.

Figure Legends

Fig. 1. PSI-BLAST and phylogenetic analyses support the HGT origin of DV-regulated ankyrin genes in chalcid wasps (CLANKs), and additional HGTs in insects. A. A portion of the phylogenetic tree generated from significant sequences using the C-terminus of *Nv-CLANK-C* as a query in PSI-BLAST. DV regulated ankyrin genes CLANK-O and -N cluster with ankyrin genes from *Wolbachia,* and other representatives of the Chalcidoidea (*Trichogramma, Ceratosolen,* and *Copidosoma*). **B.** Numerous proteins from the ant Pseudomyrmex cluster strongly and consistently with other Wolbachia ankyrin domain encoding genes. **C.** Screen capture of BLAST result illustrating the strong similarity of C-terminal sequence from a protein not annotated as having a PRANC domain (*Nv-CLANK-B*).

Fig. 2. Phylogenetic analysis of *Nasonia* DV regulated CLANKs protein family. (Left) Phylogenetic tree of CLANK proteins of interest generated using “One Click” Phylogeny Analysis (http://phylogeny.lirmm.fr/) [25]. Branch length is proportional to the number of substitutions per site. Blue text represents proteins containing a PRANC domain. (Right) Corresponding mRNA expression domain (D = dorsally, L = laterally, V = ventrally, G = no diff. expression until gastrulation, blank = ubiquitous or lack of expression), chromosomal location, number of exons (* = two differentially spliced

transcripts occur, both contain same number of exons), and number of ankyrin repeats for each CLANK protein of interest. Colors added to emphasize similar value in each column.

Fig. 3. Laterally expressed CLANKs with dynamic expansion. A1-A6 Expression of *Nv-CLANK-G* from pre-blastoderm through gastrulation. **B1-B3** Expression of *Nv-CLANK-H* from blastoderm to gastrulation. **C1-C3** Expression of *Nv-CLANK-K* from pre-blastoderm through late-blastoderm. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down (except A5, dorsal view).

Fig. 4. Laterally expressed CLANKs with dynamic retraction. A1-A6 Expression of *Nv-CLANK-A* from pre-blastoderm through gastrulation. **B1-B3** Expression of *Nv-CLANK-E* from mid to late blastoderm. **C1-C3** Expression of *Nv-CLANK-I* from blastoderm through gastrulation. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down.

Fig. 5. Dorsally expressed CLANKs. A1-A3 Expression of *Nv-CLANK-N* from early to late-blastoderm. **B1-B3** Expression of *Nv-CLANK-O* from early-blastoderm through gastrulation. **C1-C9** Expression of *Nv-CLANK-F* from pre-blastoderm through gastrulation. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down (except C7-C8, bird's eye dorsal view).

Fig. 6. Ventral and other CLANK expression patterns. A1-A3 Expression of *Nv-CLANK-M* from early-blastoderm through gastrulation. **B1-B3** Expression of *Nv-CLANK-L* from pre-blastoderm through gastrulation. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down.

Fig. 7. Distribution of pRNAi induced embryonic lethality for each CLANK of interest. Range of embryonic lethality (as a percentage) observed in clutches of embryos from pRNAi knockdown females for each CLANK and mock injected embryos. Error bars represent minimal and maximum values. Horizontal line represents median value. Red box ranges from lower to upper quartile values. t-tests were performed comparing each CLANKs lethality to mock injected lethality. Corresponding p values listed above graph (n.s. = non-significant).

Fig. 8. Effects of reducing CLANKs on *Nv-zen* expression. A1-B4 Expression of *Nv-zen* from early-blastoderm through gastrulation in control embryos. **A1-A4** Control embryos stained with DAPI to approximate embryo age. **B1-B4** *In situ* hybridization of control embryos probing for *Nv-zen* expression. Embryos (**B1-B4)** correspond to same embryos in **A1-A4**. **C1-F4** Altered expression of *Nv-zen* following pRNAi of one CLANK transcript (lower left corner) in mid-late blastoderm embryos. **C1-C4, E1-E4** Knockdown embryos stained with DAPI to approximate embryo age. **D1-D4, F1-F4** *In situ* hybridization of knockdown embryos probing for *Nv-zen* expression (“phenotype” observed, lower right). Embryos correspond to same embryos in **C1-C4, E1-E4**. All

1117 embryos are oriented with anterior to the left, posterior to the right, dorsal up, and
1118 ventral down.
1119
1120 **Fig. 9. Distribution of pRNAi phenotypes affecting *Nv-zen* expression for each**
1121 **knockdown condition.** Percentage of knockdown embryos observed with wildtype *Nv-*
1122 *zen* expression, reduced levels of *Nv-zen* expression, an incomplete or partial dorsal
1123 stripe domain of *Nv-zen*, or lacking *Nv-zen* expression completely. Mock-injected
1124 embryos were also observed for comparison and to calculate Fisher's Exact Test to
1125 determine if a significant difference ($p < 0.05$) between the two populations for the given
1126 phenotype exists (p-value < 0.05 signified by bar above graph, color corresponds to
1127 phenotype with significant difference). Schematic representation of each phenotype is
1128 shown below graph.
1129
1130 **Fig. 10. Effects of reducing CLANKs on *Nv-twi* expression. A1-B4** Wildtype
1131 expression of *Nv-twi* from early-blastoderm through gastrulation. **A1-A4** Wildtype
1132 embryos stained with DAPI to approximate embryo age. **B1-B4** *In situ* hybridization of
1133 wildtype embryos probing for *Nv-twi i* expression. Embryos (**B1-B4)** correspond to same
1134 embryos in **A1-A4**. **C1-F4** Altered expression of *Nv-twi* following pRNAi of one CLANK
1135 transcript (lower left corner) in early-blastoderm through gastrulating embryos. **C1-C4,**
1136 **E1-E4** Knockdown embryos stained with DAPI to approximate embryo age. **D1-D4, F1-**
1137 **F4** *in situ* hybridization of knockdown embryos probing for *Nv-twi* expression
1138 ("phenotype" observed, lower right). Embryos correspond to same embryos in **C1-C4,**

E1-E4. Most embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down (C4/D4, E2/F2-E4/F4 are bird's eye ventral views).

Fig. 11. Distribution of pRNAi phenotypes affecting early *Nv-twi* expression for each knockdown condition. Percentage of knockdown embryos observed with wildtype *Nv-twi* expression, a delay in the expansion of *Nv-twi* from a thin to thick ventral stripe, reduced levels of *Nv-twi* expression, or lacking *Nv-twi* expression completely. Mock-injected embryos were also observed for comparison and to calculate Fisher's Exact Test to determine if a significant difference ($p < 0.05$) between the two populations for the given phenotype exists (p-value < 0.05 signified by bar above graph, color corresponds to phenotype with significant difference). Schematic representation of each phenotype is shown below graph.

Fig. 12. Distribution of pRNAi phenotypes effecting mid-late blastoderm *Nv-twi* expression for each knockdown condition. Percentage of knockdown embryos observed with wildtype *Nv-twi* expression, a messy slug domain border of *Nv-twi*, missing/disrupted slug fork head expression, patchy slug expression, reduced levels of *Nv-twi* expression, or lacking *Nv-twi* expression completely. Mock-injected embryos were also observed for comparison and to calculate Fisher's Exact Test to determine if a significant difference ($p < 0.05$) between the two populations for the given phenotype exists (p-value < 0.05 signified by bar above graph, color corresponds to phenotype with significant difference). Schematic representation of each phenotype is shown below graph.

Fig. 13. Distribution of pRNAi phenotypes effecting late *Nv-twi* expression for each knockdown condition. Percentage of knockdown embryos observed with wildtype *Nv-twi* expression and mesodermal internalization, improper ingression of the mesoderm, or lacking *Nv-twi* expression completely. Mock-injected embryos were also observed for comparison and to calculate Fisher's Exact Test to determine if a significant difference ($p < 0.05$) between the two populations for the given phenotype exists (p-value < 0.05 signified by bar above graph, color corresponds to phenotype with significant difference). Schematic representation of each phenotype is shown below graph.

Fig. 14. *Melittobia* CLANK candidates with significant expression patterns. A1-A6 Expression of *Md-CLANK-C* from pre-blastoderm through gastrulation. **A3,A4** are dorsal views. **A3',A4'** are ventral views of the same embryo. **B1-B4** Expression of *Md-CLANK-E1* from pre-blastoderm through gastrulation. **C1-C4** Expression of *Md-CLANK-E2* from pre-blastoderm through gastrulation. **D1-D4** Expression of *Md-CLANK-F1* from pre-blastoderm through gastrulation (**D3** is a bird's eye, dorsal view). **E1-E4** Expression of *Md-CLANK-F2* from pre-blastoderm through gastrulation. **F1-F4** Expression of *Md-CLANK-G* from early blastoderm to the start of gastrulation. **F2-F4** are bird's eye ventral views. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down (unless otherwise noted).

Fig. 15. Phylogenetic analysis of *Nasonia* and *Melittobia* CLANK protein families. (Tree) Phylogenetic tree of CLANK proteins of interest. Blue = *Nasonia* CLANKs. Red =

Melittobia CLANK ortholog candidates. Black = *Nasonia* off-target sequences from *Melittobia* reciprocal BLASTs. Branch length is proportional to the number of substitutions per site [25]. (**Inset Images)** Representative images of *Melittobia* orthologs with significant RNA localization and *Nasonia* ortholog with similar pattern. Colored box highlights pairing between wasp orthologs (violet, lime, teal, tangerine). Colored lines point to orthologs phylogenetic branch on tree (Red = *Melittobia*, Blue = *Nasonia*).

Additional Files

Additional File 1. (.xlsx) *D.melanogaster* and *A.mellifera* BLASTp hits for each *Nasonia* novel ankyrin-repeat containing transcript. Each of the 15 novel *Nasonia* transcripts was entered into a BLASTp query to search for homologous proteins in *Drosophila* and *Apis*. The top 100 hits were collected (duplicate sequences were removed) for each *Nasonia* transcript (individual sheets within workbook). Each row represents a distinct BLASTp hit and contains the hit's description (column A), BLAST e-value (column B), and NCBI accession number (column C).

Additional File 2. (.xlsx) Top *D.melanogaster* and *A.mellifera* BLASTp novel ankyrin-repeat hits. All sequences from Additional file 1 were pooled (duplicate sequences removed) and entered as a reciprocal BLASTp query to see if they were strongly homologous to the *Nasonia* novel ankyrin-repeat containing transcripts. Each row represents a distinct BLASTp hit from Additioanl file 1, and again contains the hit's description (column A, Yellow highlights *Apis* sequences. *Drosophila* sequences are left

in white), BLAST e-value (column B), and NCBI accession number (column C) as well as the description of the *Nasonia* protein it reciprocal BLASTs back to (column E), BLAST e-value (column F), and NCBI accession number (column G).

Additional File 3. (.docx) Phylogenetic Analyses of the top 100 blast hits for each of the 15 Nv-CLANKs analyzed in this paper. CLANK sequences are in red, and bacterial and viral sequences are in green. Clades containing many Trichomonas ankyrin domain encoding genes have been collapsed.

Additional File 4. (.pdf) Phylogenetic analysis of Psi-BLAST of the C-termini of Nv-CLANKs. CLANK sequences are in red, and bacterial and viral sequences are in green. Large clades containing sequences exclusively from a single species have been collapsed. See label for panel number and CLANK used as query to generate the tree.

Additional File 5. (.xlsx) Relationships between our sequence nomenclature and gene identification numbers in different annotations. Sheet 1. Reference sequences and names corresponding to each novel ankyrin-repeat containing *Nasonia* transcripts. "Working Code", "Transcript" number from annotation 2.0 of the *N. vitripennis* genome, and corresponding "Gene Symbol", "NCBI Reference Sequence" (mRNA), and "NCBI Reference Sequence" (protein) for each transcript of interest. Alternatively, spliced transcripts and proteins are listed as a second row for a given "Transcript." **Sheet 2. CLANK ortholog candidate sequences in *Melittobia***. Column A: Working code used throughout paper to identify sequences. Column B: *Melittobia*

CLANK sequence accession numbers from de novo embryonic transcriptome (in prep). **Sheet 3. *Nasonia* off-target hits from *Melittobia* CLANK ortholog BLAST**. Column A: Working code used throughout paper to identify sequences. Column B: *Nasonia* sequence accession numbers from annotation 2.0 of the *N. vitripennis* genome. Column C: *Nasonia* NCBI Reference Sequences. Column D: *Nasonia* NCBI Reference Sequences predicted gene name. **Sheet 4. *Melittobia* primer sequences for *in situ* hybridization probes**. Column A: Primer name. Column B: Primer sequence. Primers were designed using Primer3 v.0.4.0 (**http://primer3.ut.ee)** and synthesized by Integrated DNA Technologies (IDT, www.idtdna.com/Site/Order/oligoentry).

Additional File 6. (.tiff) CLANKs lacking differential expression patterns. A1-D3 Expression of *Nv-CLANK-B*, *-C*, *-D*, *-J* from pre-blastoderm through gastrulation. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down.

Additional file 7. (.tiff) Distribution of pRNAi survival rate for each CLANK of interest. Range of pupal survival and eclosure (as a percentage) observed in pRNAi knockdown females for each CLANK and mock injection. Error bars represent minimal and maximum values. Horizontal line represents median value. Red box ranges from lower to upper quartile values.

Additional File 8. (.tiff) Relative embryonic expression of CLANK transcripts over time following pRNAi. cDNA was generated from aged (3-7 h, 28^{o}C) embryos,

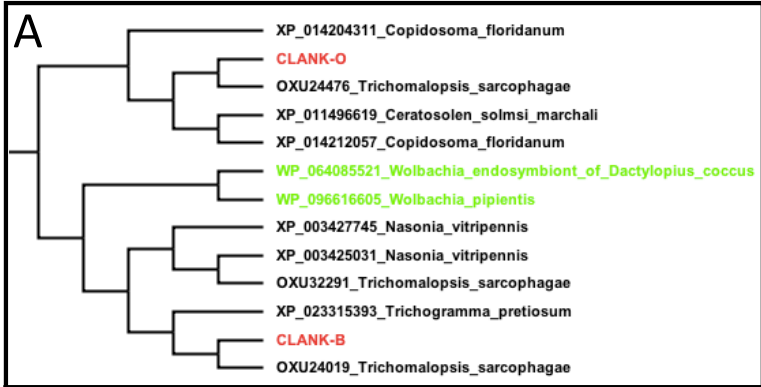
collected from pRNAi injected females for up to three days post eclosure. mRNA expression levels of the knockdown transcript were monitored via qPCR. Relative expression compared to mock injected embryos (as a percentage out of 100) was calculated and plotted after reactions were normalized via *Nv-rp49* expression. Expression values are an average of biological and technical replicates.

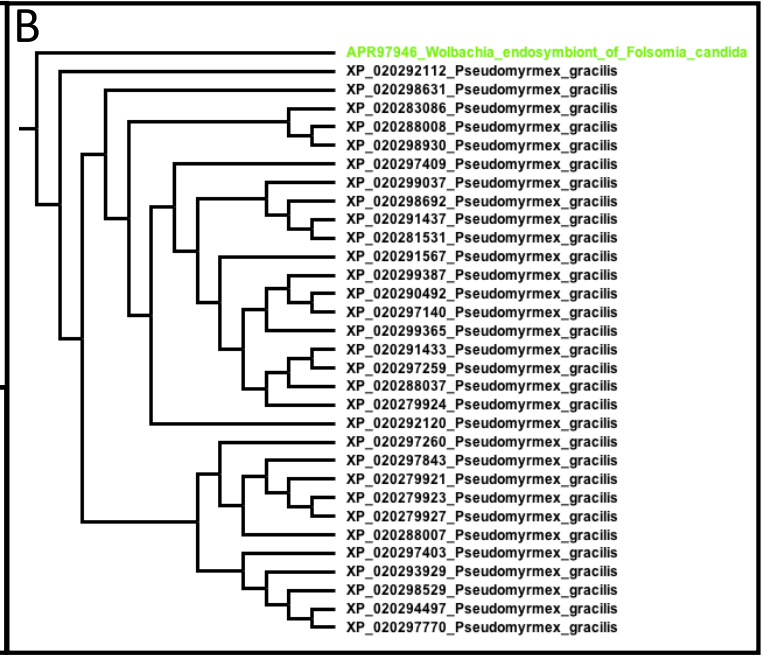
Additional File 9. (.tiff) Effects of reducing *CLANKs* on *Nv-zen* expression. **A-U'** Altered expression of *Nv-zen* following pRNAi of a *CLANK* of interest in mid-late blastoderm embryos. **A-U** Knockdown embryos stained with DAPI to approximate embryo age. **A'-U'** *In situ* hybridization of knockdown embryos probing for *Nv-zen* expression. Embryos correspond to same embryos in **A-U**. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down. **A-B'** *Nv-CLANK-A* pRNAi embryos. **C-D'** *Nv-CLANK-G* pRNAi embryos. **E-G'** *Nv-CLANK-E* pRNAi embryos. **H-K'** *Nv-CLANK-F* pRNAi embryos. **L-M'** *Nv-CLANK-H* pRNAi embryos. **N-O'** *Nv-CLANK-I* pRNAi embryos. **P-R'** *Nv-CLANK-N* pRNAi embryos. **S-U'** *Nv-CLANK-O* pRNAi embryos. Descriptive term of phenotype observed in bottom right corner of *in situ* images.

Additional File 10. (.tiff) Effects of reducing *CLANKs* on *Nv-twi* expression. **A-EE'** Altered expression of *Nv-twi* following pRNAi of a *CLANK* of interest in mid blastoderm to gastrulating embryos. **A-EE** Knockdown embryos stained with DAPI to approximate embryo age. **A'-EE'** *In situ* hybridization of knockdown embryos probing for *Nv-twi* expression. Embryos correspond to same embryos in **A-EE**. All embryos are oriented

with anterior to the left, posterior to the right, dorsal up, and ventral down (unless otherwise noted). **A-D'** *Nv-CLANK-A* pRNAi embryos (C/C', D/D' ventral views). **E-F'** *Nv-CLANK-E* pRNAi embryos. **G-I'** *Nv-CLANK-H* pRNAi embryos (H/H' ventral views). **J-Q'** *Nv-CLANK-G* pRNAi embryos (K-N' ventral views). **R-U'** *Nv-CLANK-I* pRNAi embryos (S/S', U/U' ventral views). **V-Z'** *Nv-CLANK-K* pRNAi embryos. **AA-EE'** *Nv-CLANK-M* pRNAi embryos (AA-EE' ventral views). Descriptive term of phenotype observed in bottom right corner of *in situ* images.

Additional File 11. (.tiff) ***Melittobia*** **CLANK candidates lacking differential expression patterns. A1-J3** Expression of *Md-CLANK-A*, *-B*, *-D*, *-H*, *-I1, -I2*, *-J*, *-K*, *-L*, and *-N* from pre-blastoderm through gastrulation. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down (except B3, bird's eye dorsal view).





C

MULTISPECIES: PRANC domain-containing protein [Wolbachia]

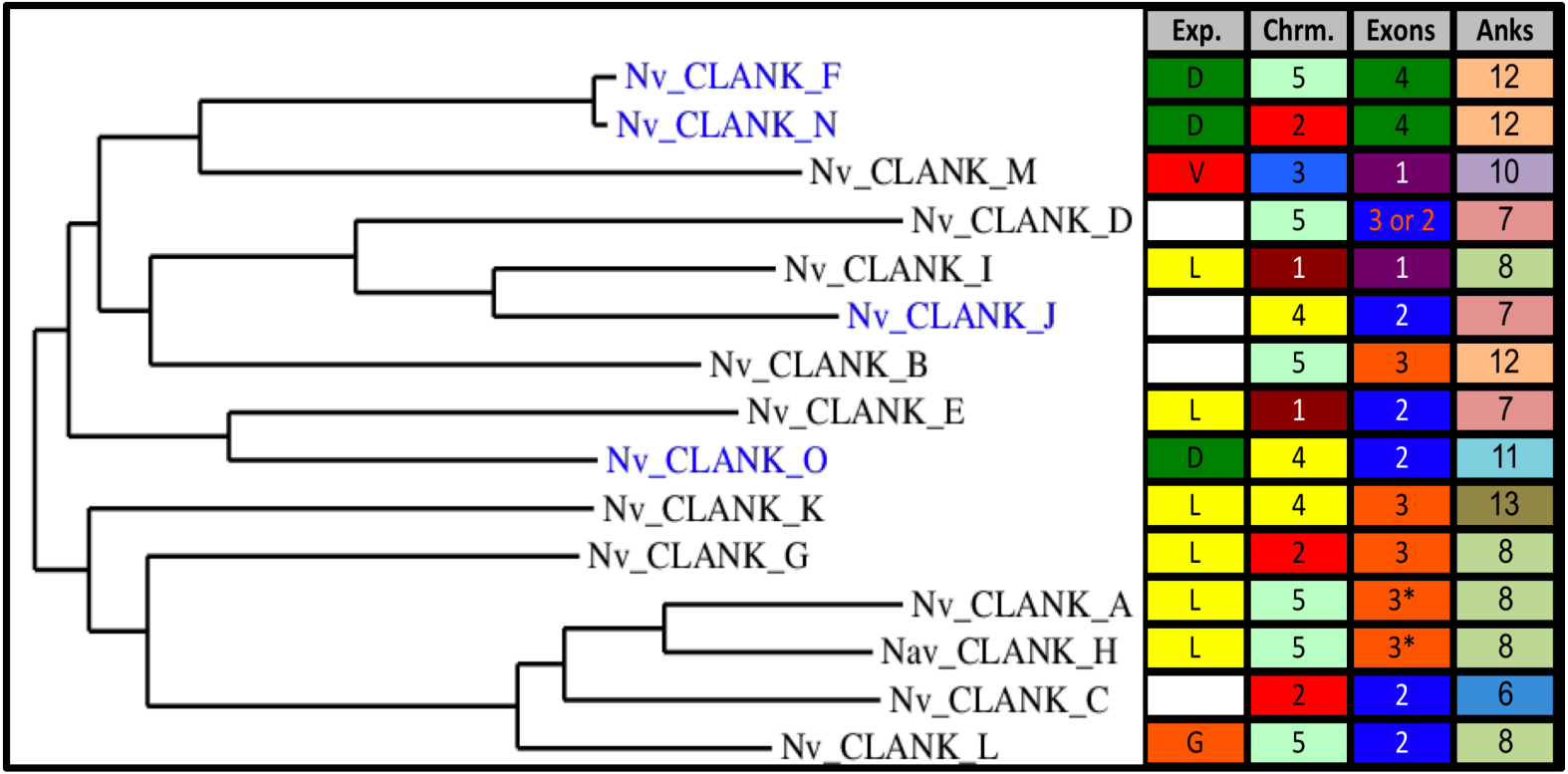
Sequence ID: WP_010405507.1 Length: 946 Number of Matches: 1

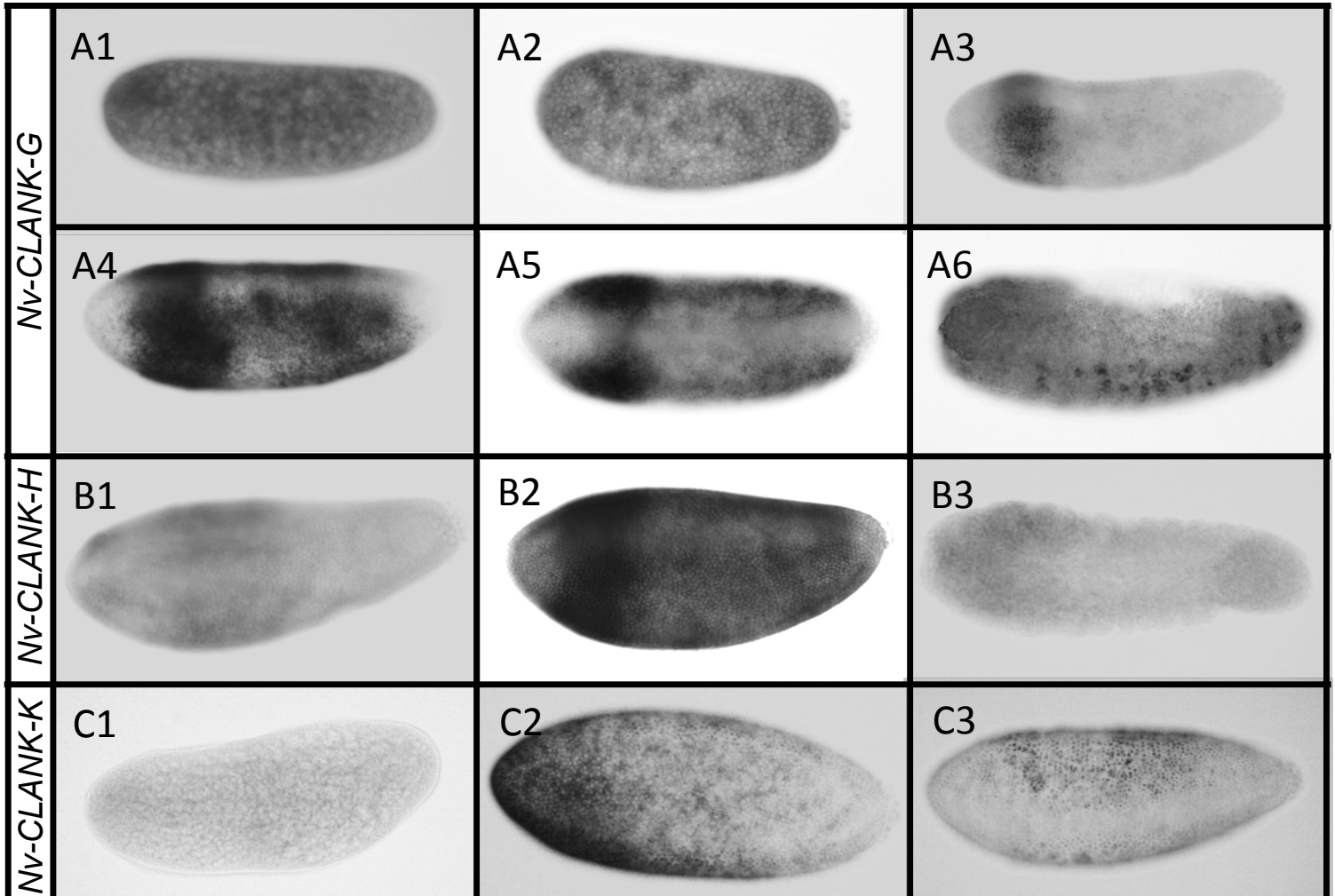
▸See 4 more title(s)

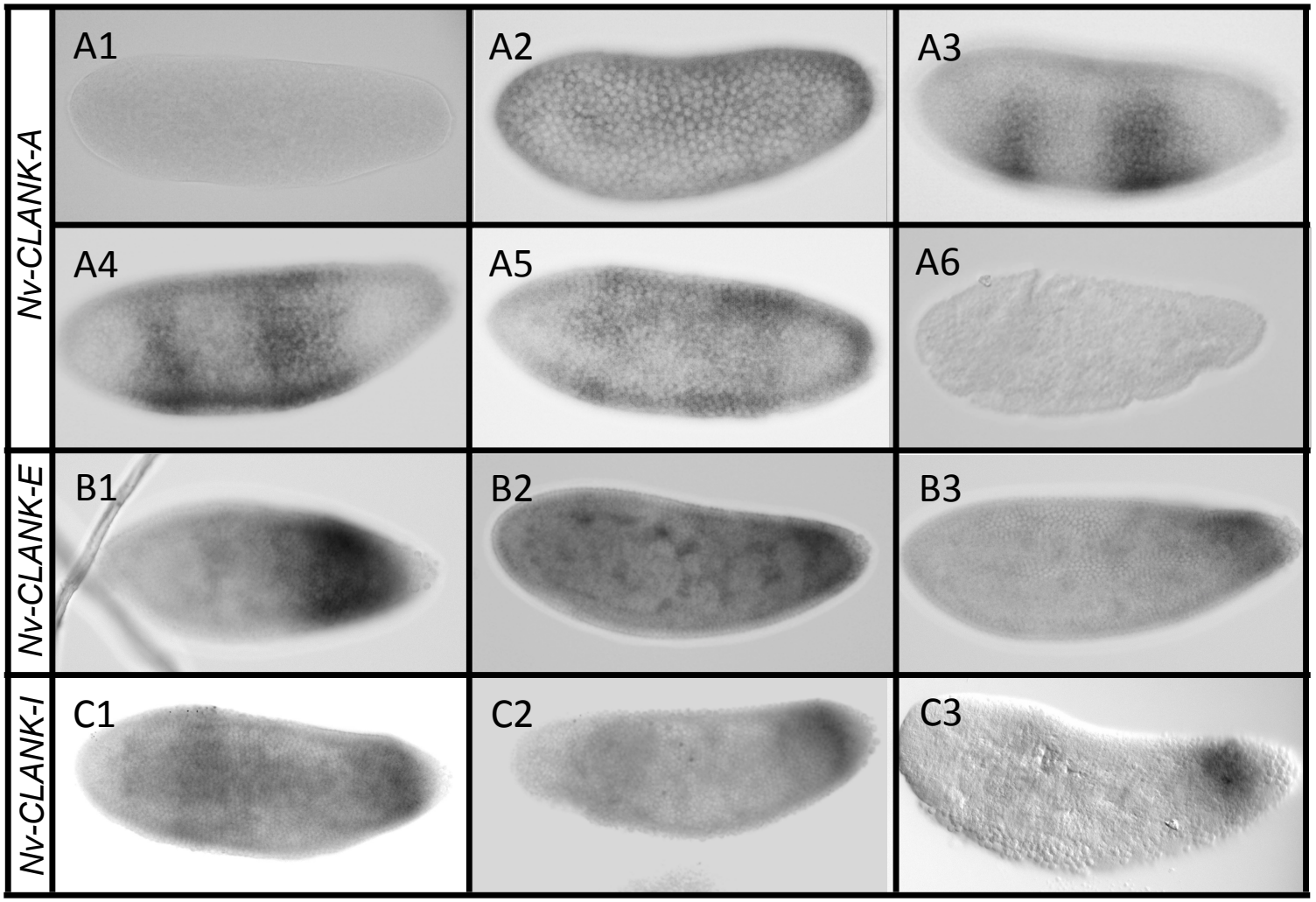
Range 1: 770 to 920 GenPept Graphics ▼ Next Match ▲ Previous

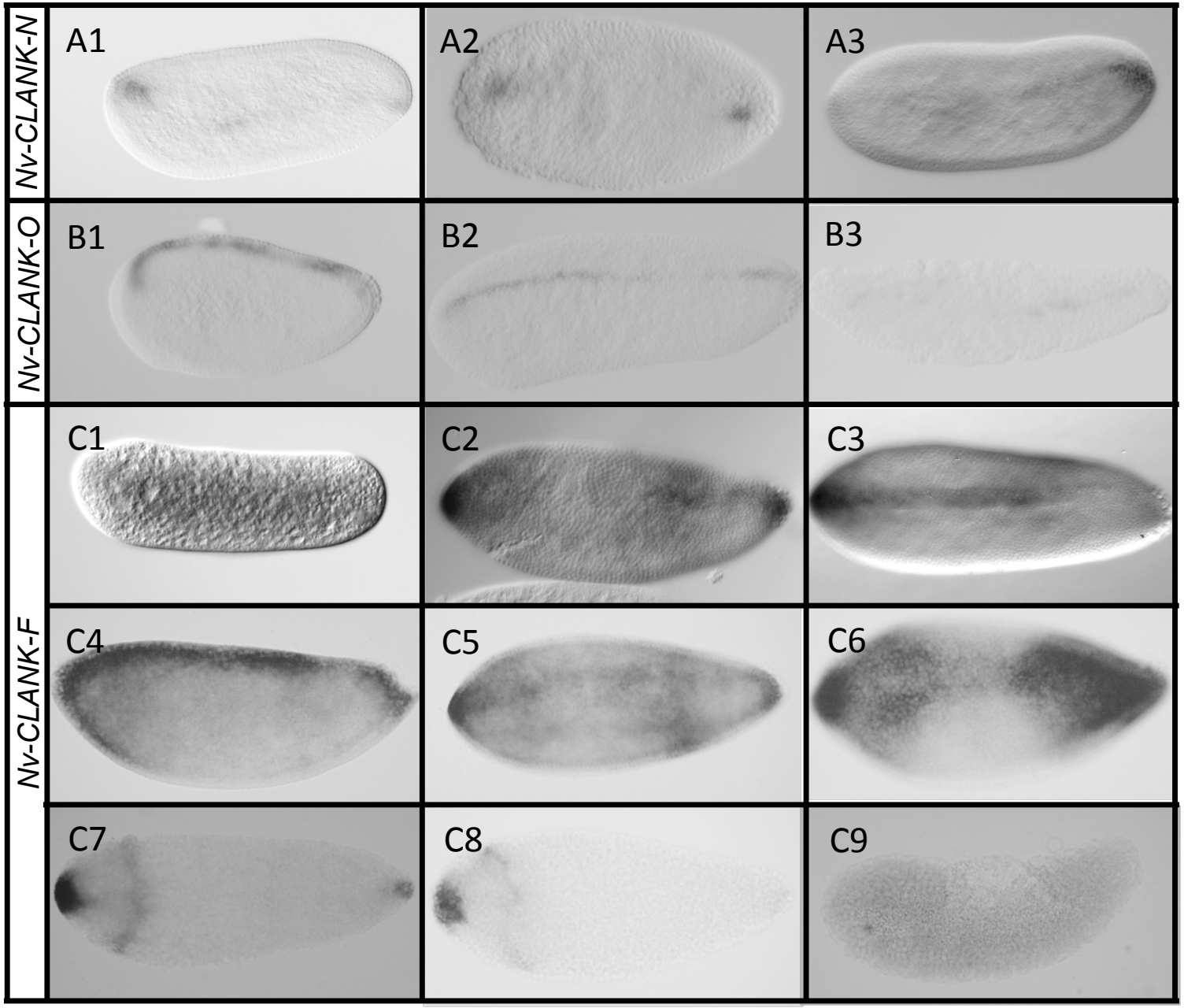
Score	Expect	Method	Identities	Positives	Gaps
137 bits(346)	5e-34	Composition-based stats.	34/152(22%)	64/152(42%)	3/152(1%)

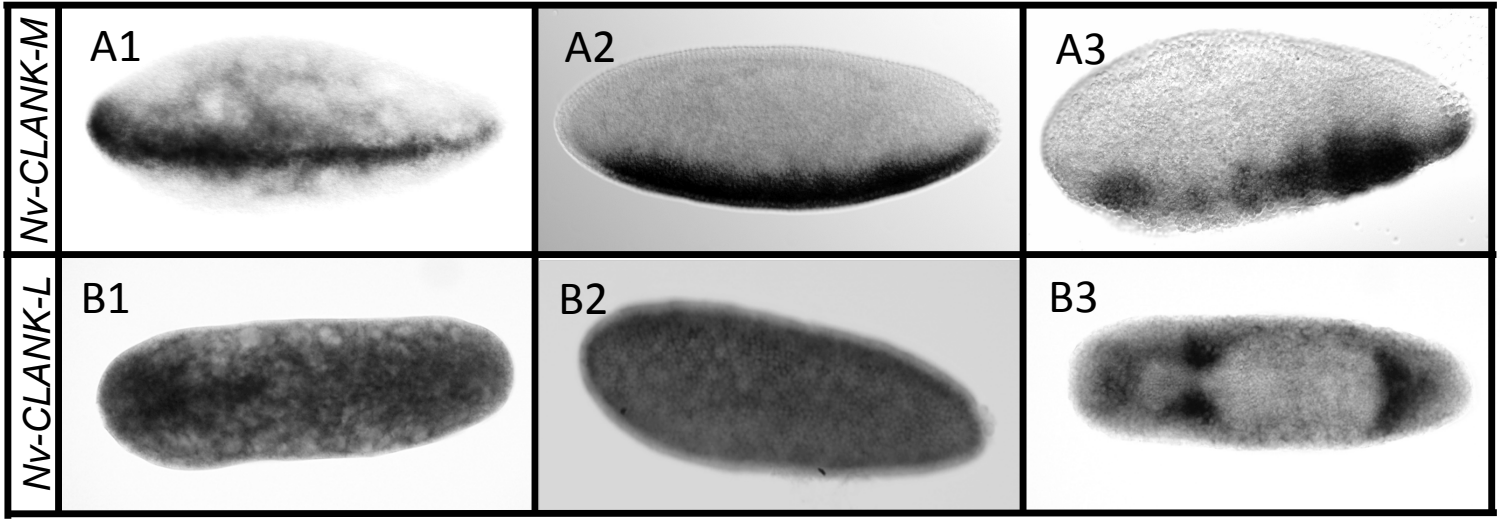
```
Query  25   LVRHMALMTVRQQFIGVKNVDAILSRPEYFTIFQLSLREIHLMQKTKIGNSNVFYYDLLT  84
            L RH+  M       E      KN+ ++     EI  M+  KI N N+ +Y +L
Sbjct  770  LKRHIVKMKTANLYVSEKNLLSMSISGEISGFQDECEEEIARMKSEKINNPNISFYGILA  829

Query  85   GKGRQLVKYAQNVDVAETMEAQKYRLAYCHYYCMQLIVNYTTCKTRLNLIRQASIKLVEL  144
                 L  Y ++ ++ + +++  Y+  +  Y  M +  ++  G  R  L+ Q +     L
Sbjct  830  KGTSSLAIYMRDENIVQVLKSDDYKTKFPIYASM-IESHFRKGMERKELLEQGNKIFHFL  888

Query  145  APV--KLPTEIATYICEHLSNKNLLNLEKTKL  174
                 +LP +   I   +LSN+ +L  L
Sbjct  889  FNNFPELPYDCTEKIFSYLSNEDLRILIDACK  920
```

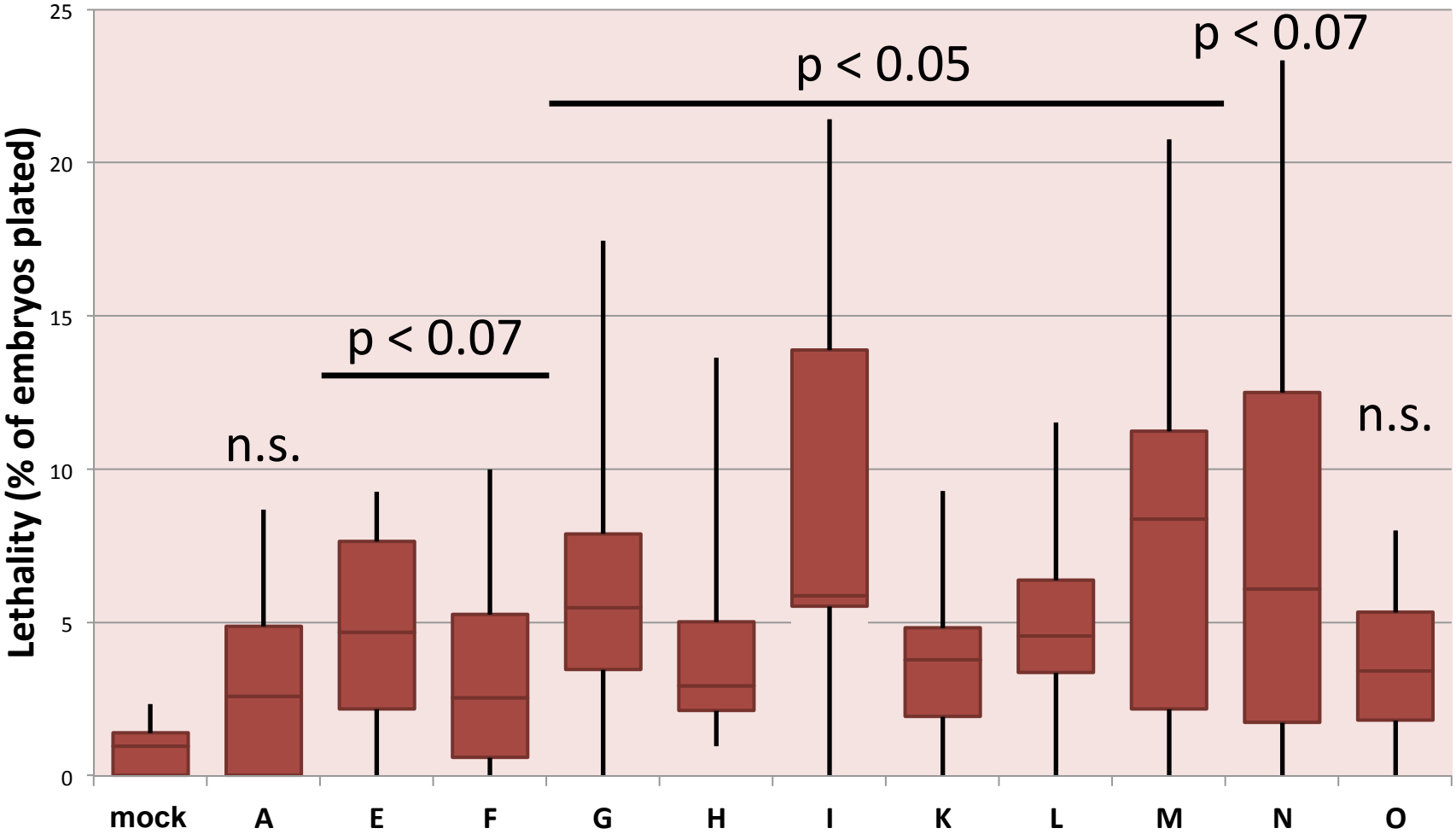

Nv_CLANK_F
Nv_CLANK_N
Nv_CLANK_M
Nv_CLANK_D
Nv_CLANK_I
Nv_CLANK_J
Nv_CLANK_B
Nv_CLANK_E
Nv_CLANK_O
Nv_CLANK_K
Nv_CLANK_G
Nv_CLANK_A
Nav_CLANK_H
Nv_CLANK_C
Nv_CLANK_L
Exp.
Chrm.
Exons
Anks
D 5 4 12
D 2 4 12
V 3 1 10
5 3 or 2 7
L 1 1 8
4 2 7
5 3 12
L 1 2 7
D 4 2 11
L 4 3 13
L 2 3 8
L 5 3* 8
L 5 3* 8
2 2 6
G 5 2 8
0.7

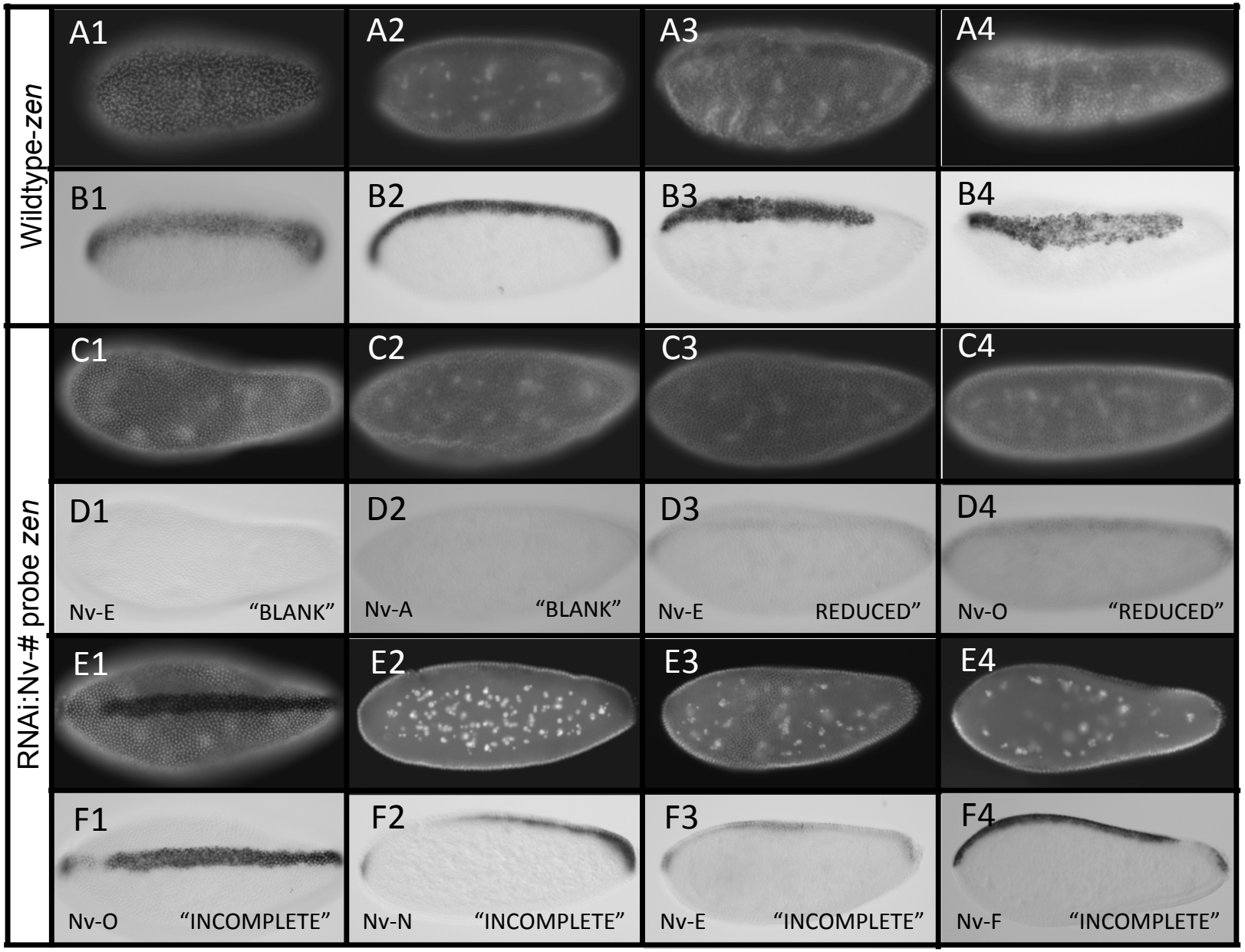

Nv-CLANK-G
A1
A2
A3
A4
A5
A6
Nv-CLANK-H
B1
B2
B3
Nv-CLANK-K
C1
C2
C3

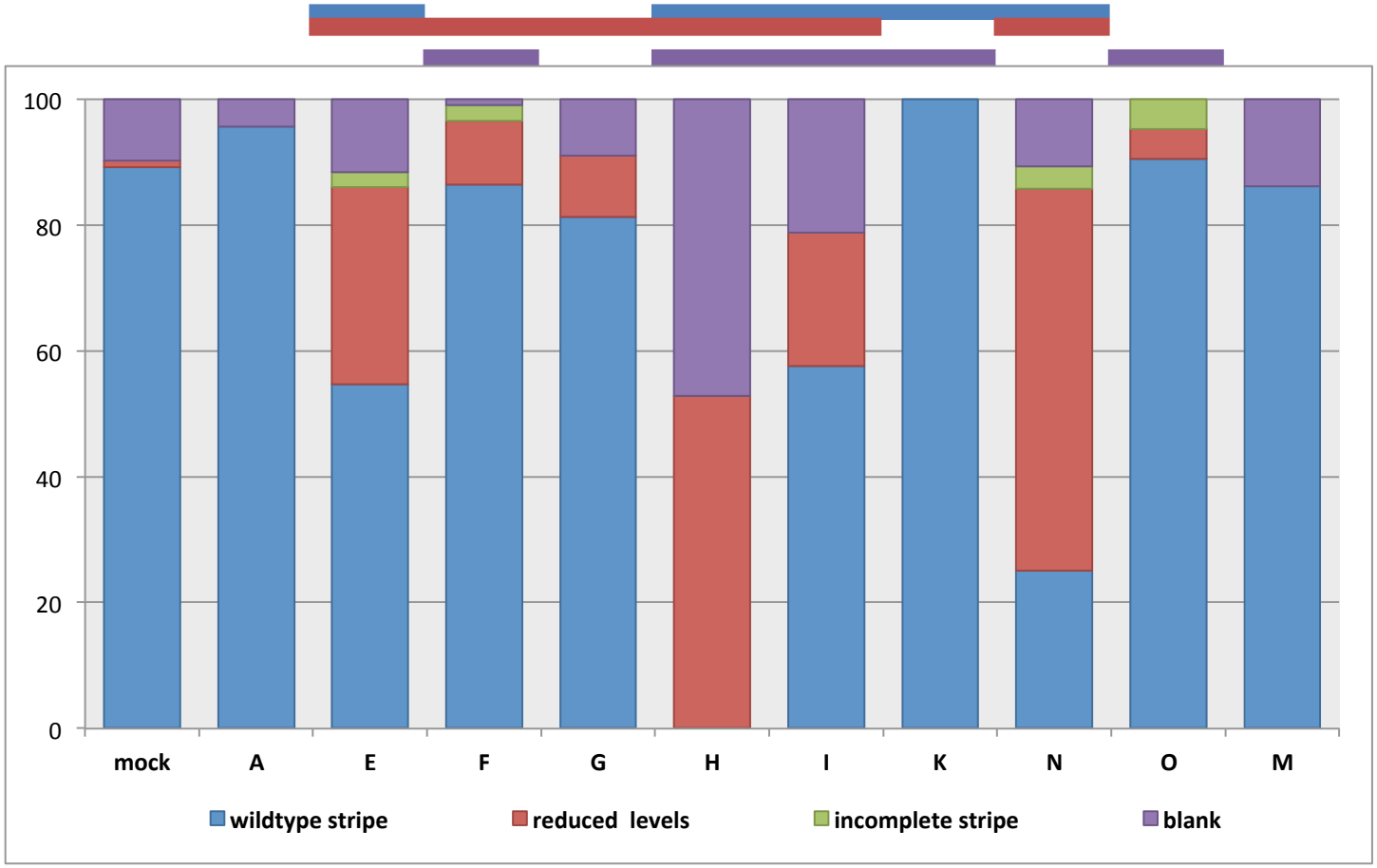

Nv-CLANK-A
A1
A2
A3
A4
A5
A6
Nv-CLANK-E
B1
B2
B3
Nv-CLANK-I
C1
C2
C3

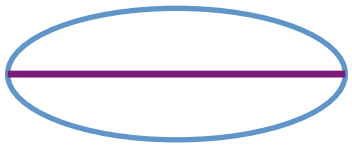

Nv-CLANK-N
A1
A2
A3
Nv-CLANK-O
B1
B2
B3
Nv-CLANK-F
C1
C2
C3
C4
C5
C6
C7
C8
C9

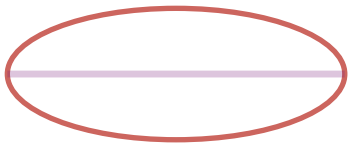

Nv-CLANK-M
A1
A2
A3
Nv-CLANK-L
B1
B2
B3

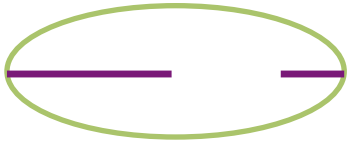

Lethality (% of embryos plated)
25
20
15
10
5
0
p < 0.05
p < 0.07
p < 0.07
n.s.
n.s.
mock
A
E
F
G
H
I
K
L
M
N
O

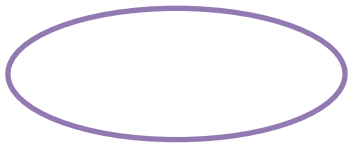

Wildtype-*zen*
RNAi:Nv-# probe *zen*
A1
A2
A3
A4
B1
B2
B3
B4
C1
C2
C3
C4
D1
Nv-E
"BLANK"
D2
Nv-A
"BLANK"
D3
Nv-E
REDUCED"
D4
Nv-O
"REDUCED"
E1
E2
E3
E4
F1
Nv-O
"INCOMPLETE"
F2
Nv-N
"INCOMPLETE"
F3
Nv-E
"INCOMPLETE"
F4
Nv-F
"INCOMPLETE"

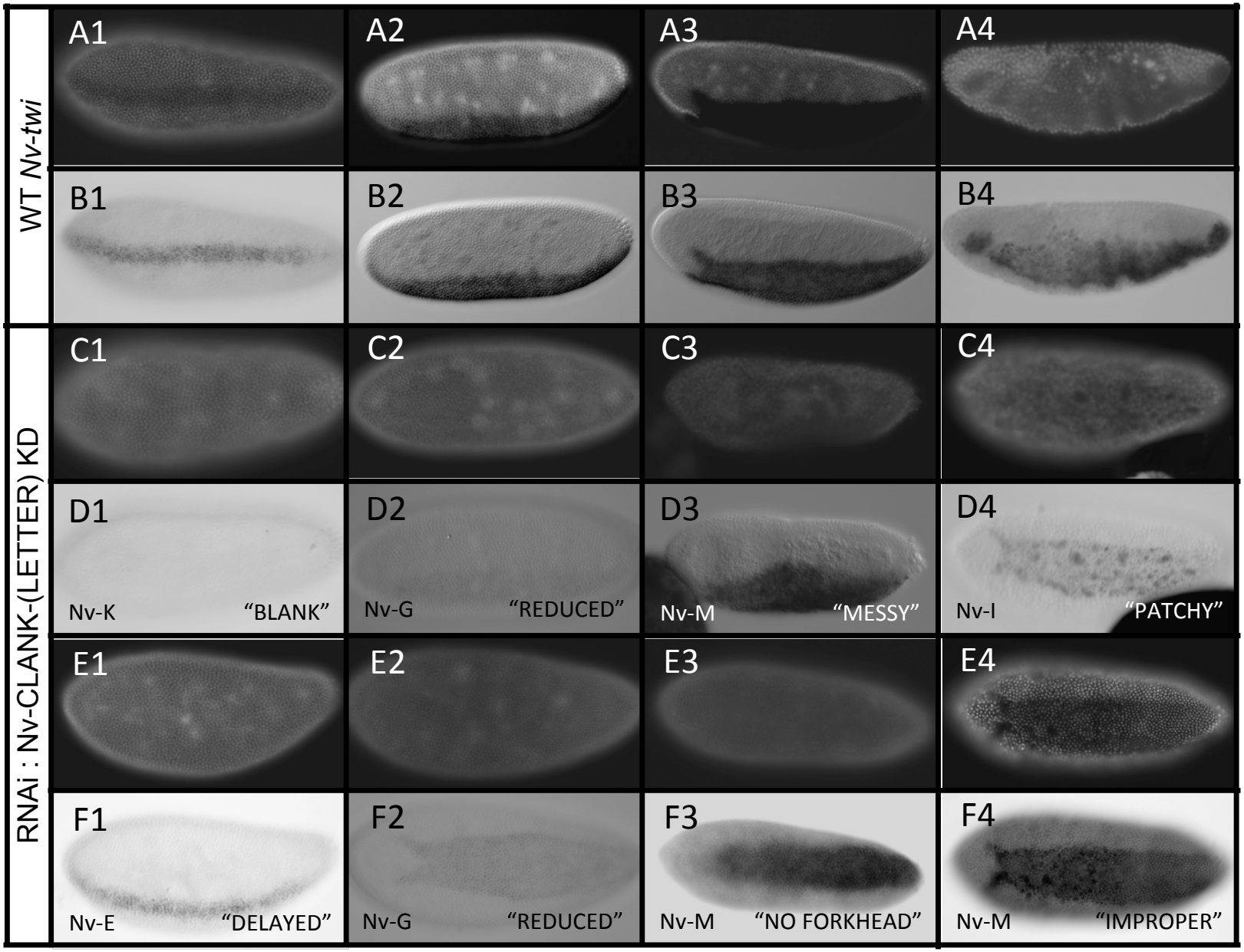

100
80
60
40
20
0
mock
A
E
F
G
H
I
K
N
O
M
wildtype stripe
reduced levels
incomplete stripe
blank

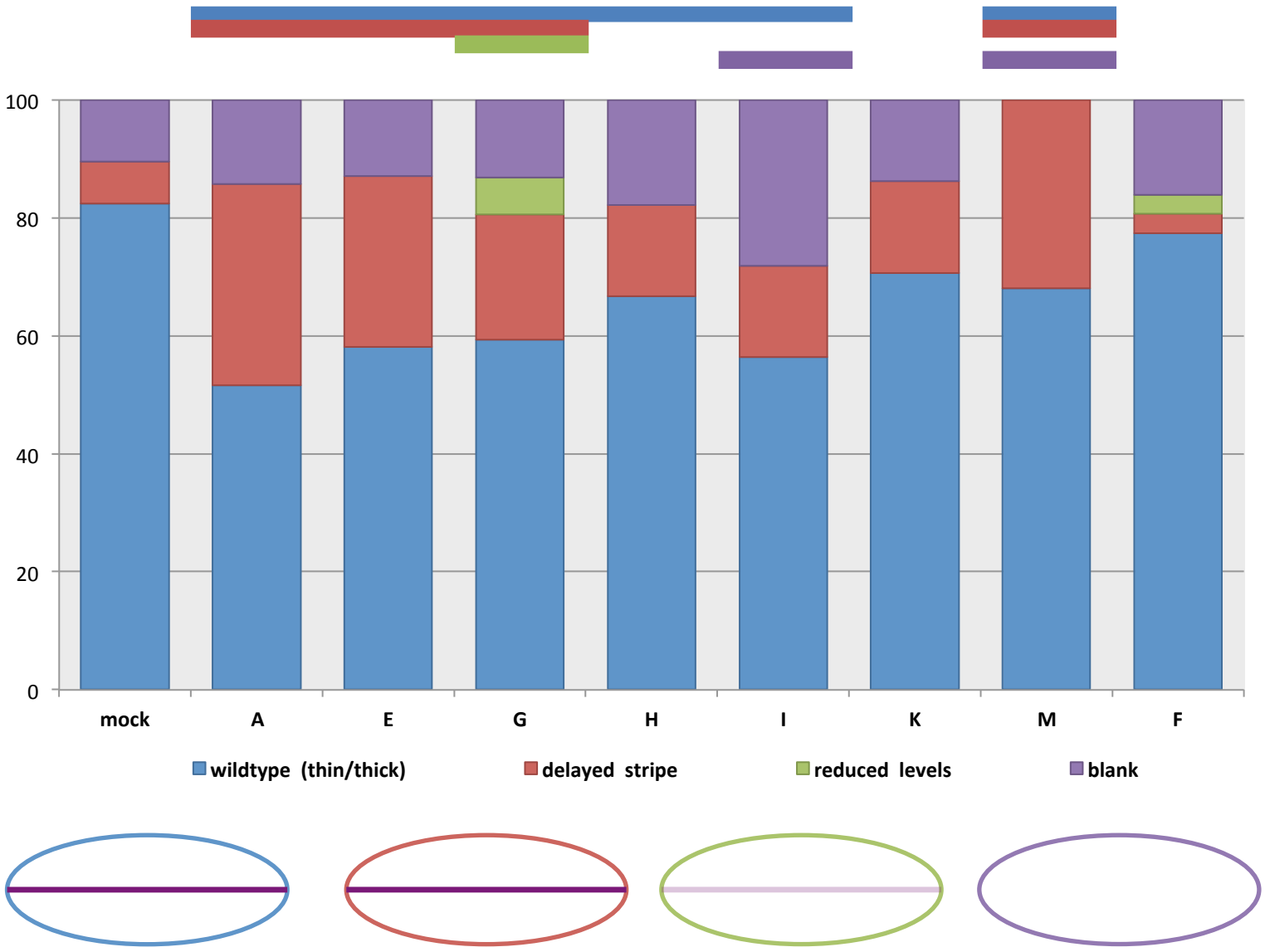


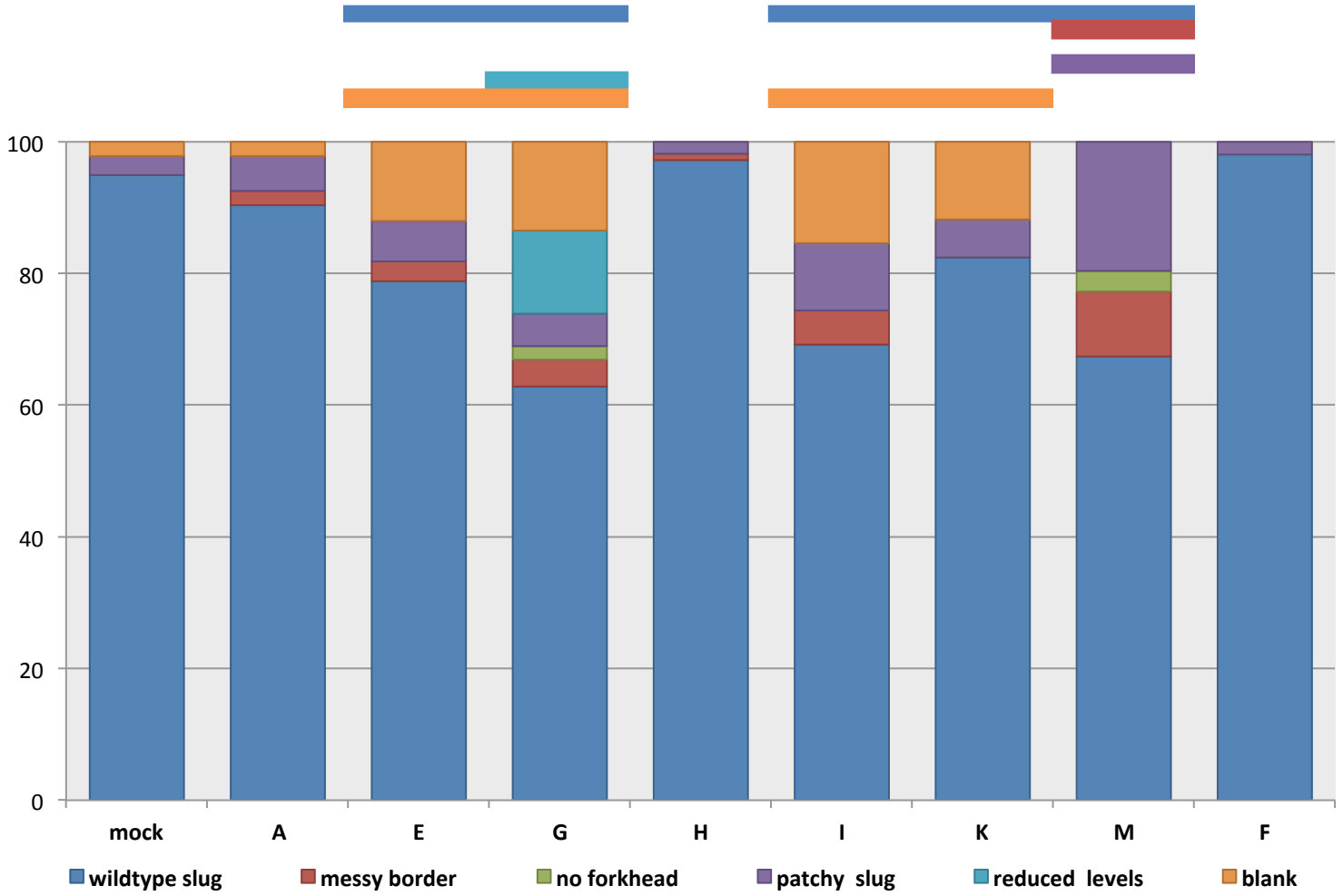


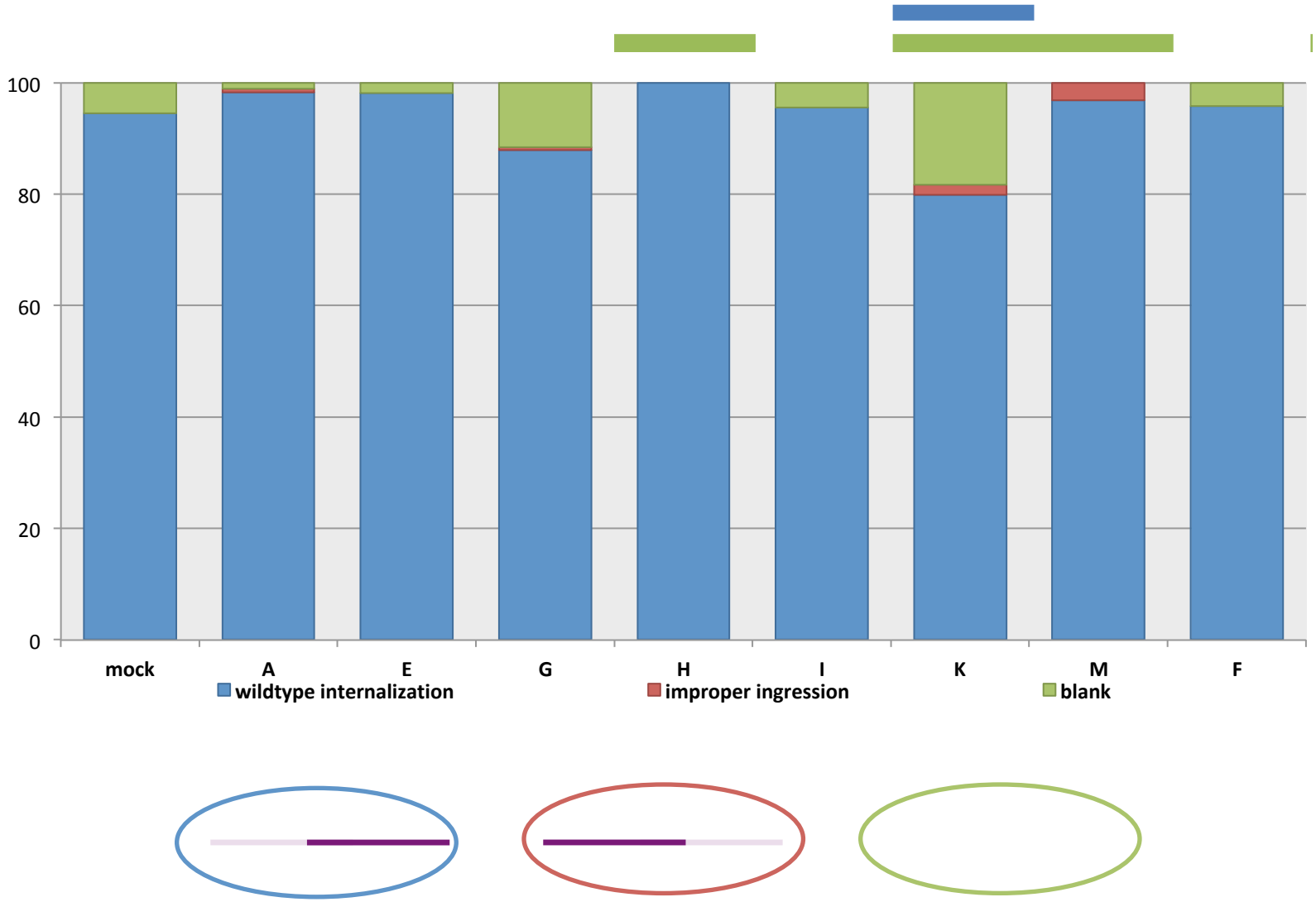


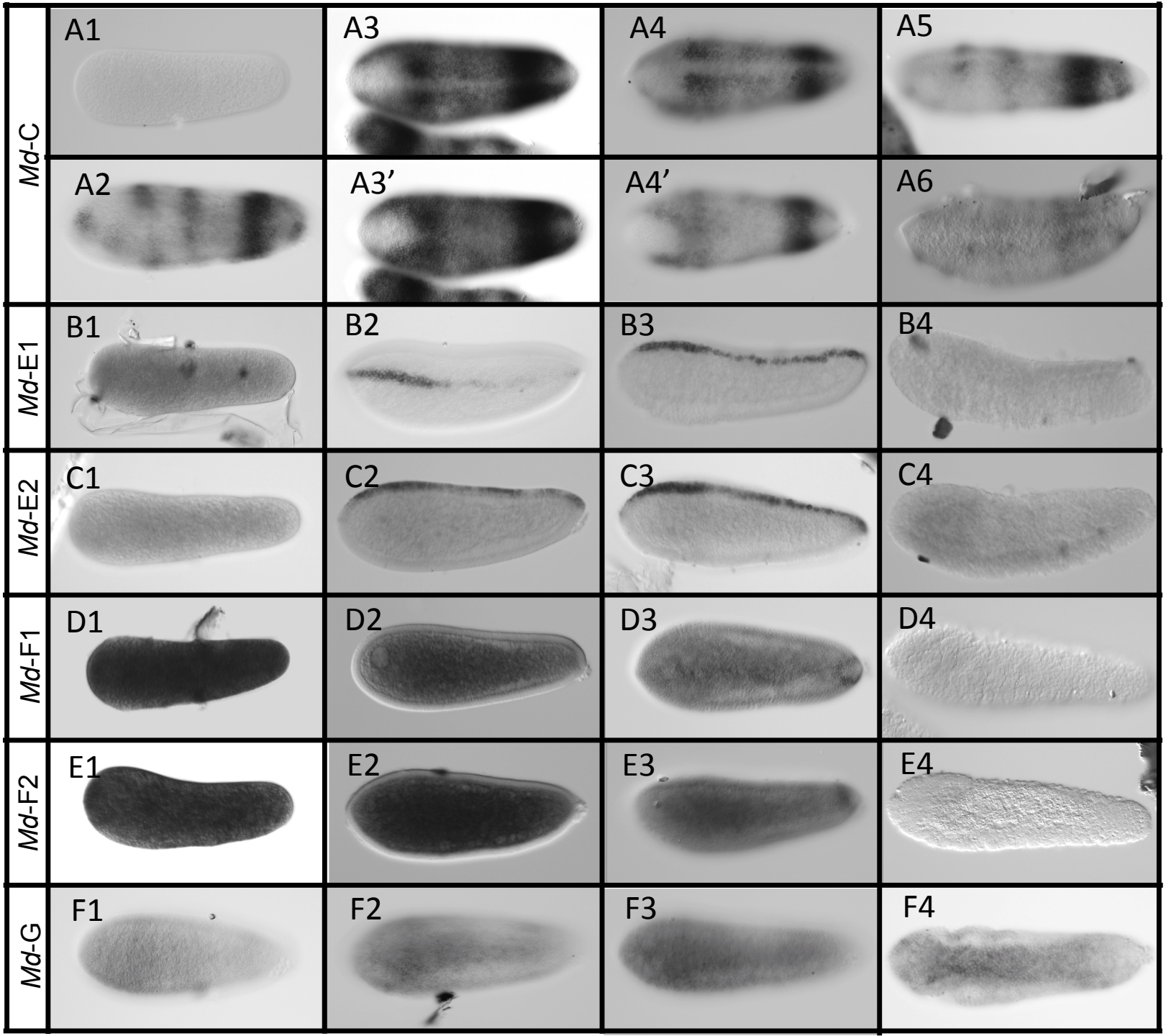


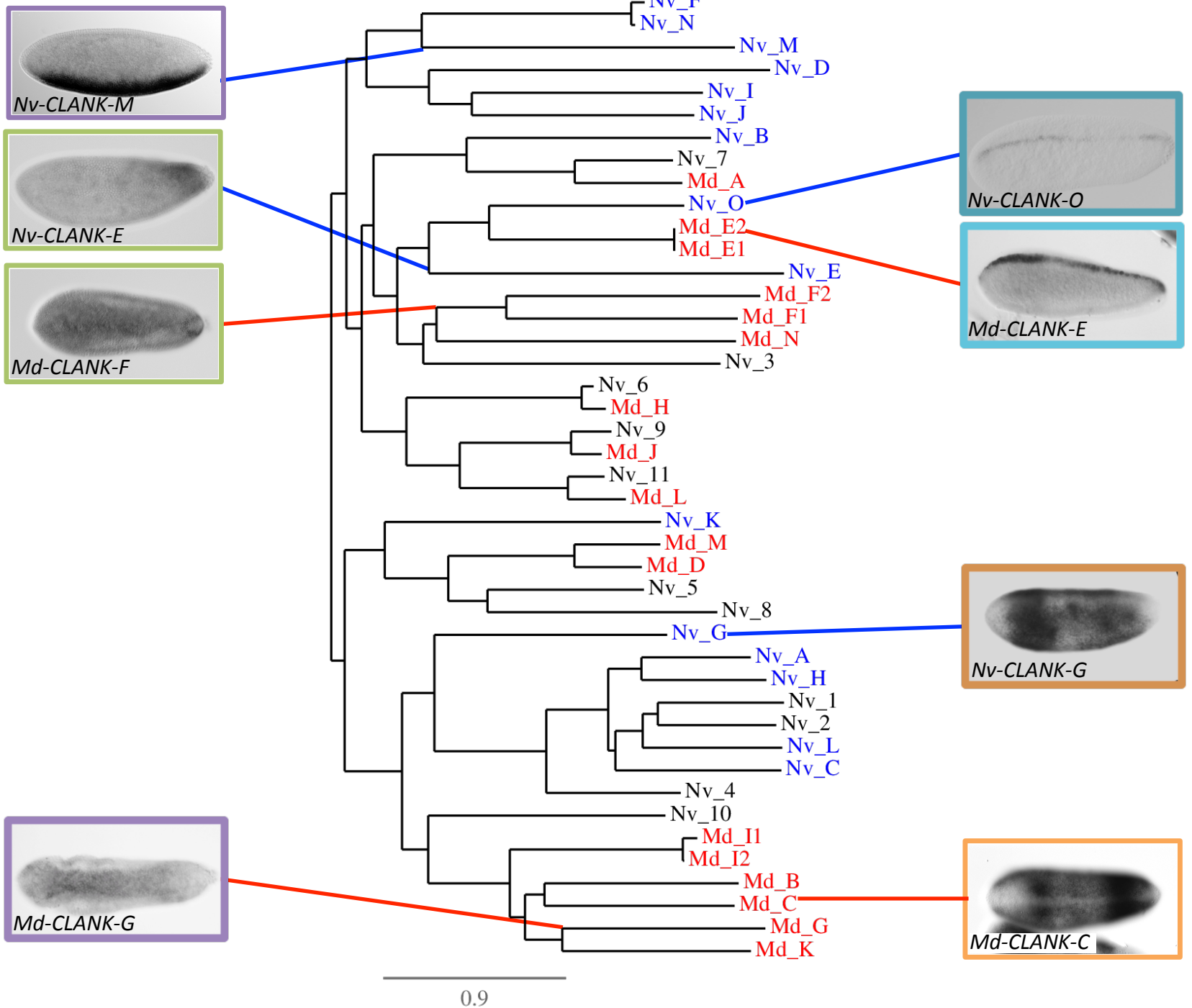

WT Nv-twi
RNAi : Nv-CLANK-(LETTER) KD
A1
A2
A3
A4
B1
B2
B3
B4
C1
C2
C3
C4
D1
Nv-K
"BLANK"
D2
Nv-G
"REDUCED"
D3
Nv-M
"MESSY"
D4
Nv-I
"PATCHY"
E1
E2
E3
E4
F1
Nv-E
"DELAYED"
F2
Nv-G
"REDUCED"
F3
Nv-M
"NO FORKHEAD"
F4
Nv-M
"IMPROPER"


100
80
60
40
20
0
mock
A
E
G
H
I
K
M
F
wildtype (thin/thick)
delayed stripe
reduced levels
blank


100
80
60
40
20
0
mock
A
E
G
H
I
K
M
F
wildtype slug
messy border
no forkhead
patchy slug
reduced levels
blank


100
80
60
40
20
0
mock
A
E
G
H
I
K
M
F
wildtype internalization
improper ingression
blank


Md-C
A1
A3
A4
A5
A2
A3’
A4’
A6
Md-E1
B1
B2
B3
B4
Md-E2
C1
C2
C3
C4
Md-F1
D1
D2
D3
D4
Md-F2
E1
E2
E3
E4
Md-G
F1
F2
F3
F4


Nv_F
Nv_N
Nv_M
Nv_D
Nv_I
Nv_J
Nv_B
Nv_7
Md_A
Nv_O
Md_E2
Md_E1
Nv_E
Md_F2
Md_F1
Md_N
Nv_3
Nv_6
Md_H
Nv_9
Md_J
Nv_11
Md_L
Nv_K
Md_M
Md_D
Nv_5
Nv_8
Nv_G
Nv_A
Nv_H
Nv_1
Nv_2
Nv_L
Nv_C
Nv_4
Nv_10
Md_I1
Md_I2
Md_B
Md_C
Md_G
Md_K
Nv-CLANK-M
Nv-CLANK-E
Md-CLANK-F
Nv-CLANK-O
Md-CLANK-E
Nv-CLANK-G
Md-CLANK-G
Md-CLANK-C
0.9