RESEARCH

Open Access

CrossMark

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

Daniel Pers and Jeremy A. Lynch*

Abstract

Background: How regulatory networks incorporate additional components and how novel genes are functionally integrated into well-established developmental processes are two important and intertwined questions whose answers have major implications for understanding the evolution of development. We recently discovered a set of lineage-restricted genes with strong and specific expression patterns along the dorsal-ventral (DV) axis of the embryo of the wasp *Nasonia* that may serve as a powerful system for addressing these questions. We sought to both understand the evolutionary history of these genes and to determine their functions in the *Nasonia* DV patterning system.

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 a prokaryote in a common ancestor of the wasp superfamily 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.

Conclusions: Our results indicate that regulatory networks can incorporate novel genes that then become necessary for stable and repeatable outputs. Even a modest role in developmental networks may be enough to allow novel or duplicate genes to be maintained in the genome and become fully integrated network components.

Keywords: Regulatory networks, Development, EvoDevo, Ankyrin, Chalcidoidea, *Nasonia*, Dorsoventral patterning, Embryo, *Wolbachia*, Gene duplication, Horizontal gene transfer

Background

Gene regulatory networks (GRNs) coordinate the expression of mRNA and proteins in a spatiotemporal manner to bring about a specific developmental output [1]. The complex webs of interacting nodes and modules that make up GRNs are vital for establishing patterning, morphogenesis, and ultimately an organism's body plan [2]. Perturbations to these networks should result in novel developmental outputs. However, canalization and developmental redundancy can conceal underlying genetic variation and phenotypic plasticity. This quality enables GRNs to weather large variations in genomic and environmental inputs, without disrupting the phenotypic output of the network [3–6].

These properties of GRNs raise questions about how developmental mechanisms can evolve. Since robust networks can absorb large genetic changes without causing major changes in developmental output, it seems that a large threshold must be overcome in order to achieve a new phenotype [7]. Thus, robustness paradoxically may make GRNs less able to respond to evolutionary pressures since most mutations will not produce phenotypes visible to natural selection. Thus, we might expect

* Correspondence: jlynch42@uic.edu
Department of Biological Sciences, University of Illinois at Chicago, MBRB
4020, 900 S. Ashland Ave, Chicago, IL 60607, USA

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* (*N. vitripennis*) has been shown to be a good model to study novelty and the evolution of gene networks. Having split from *Drosophila melanogaster* (*D. melanogaster*) over 300 MYA [11], *Nasonia* has converged on a similar mode of embryogenesis [12] and shares 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 evidence that these genes entered the genome of the ancestor of *Nasonia* through at least one horizontal gene transfer (HGT) event, followed by several waves of duplication and divergence. We have previously shown that these ankyrin domain-encoding genes are expressed in specific patterns along the DV axis [15], and here, we demonstrate that they also are functionally incorporated into the DV patterning GRN, as their loss leads to variable disruptions to patterning. Through examination of another wasp, *Melittobia digitata (M. digitata)*, we also show that some of the functional incorporation is ancient within the Superfamily, while there is also strong evidence of recent gains and/or losses of function in the *Nasonia* and *Melittobia* lineages.

We propose that the properties of ankyrin domain-containing proteins allow them to rapidly gain interaction partners, and potentially adaptive functions in developmental networks, which increase the likelihood that genes of this type will be maintained and sometimes multiply in the course of genome evolution.

Results

Identification of new families of apparently horizontally transferred, ankyrin domain-encoding genes

In our previous study, we identified 15 transcripts encoding ankyrin domain proteins that appeared to be significantly regulated by the Toll and/or BMP signaling pathways in the *Nasonia* embryo [15]. Further analysis of their expression showed that 6 of these genes are expressed laterally, 3 are expressed on the dorsal surface of the embryo, 1 is expressed over the ventral midline, 1 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 4 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* [17].

While the remaining 11 DV-regulated ankyrin domain-encoding genes are not annotated with PRANC domains, we believe that they entered the *Nasonia* genome through a process of horizontal transfer, gene duplication, and rapid molecular divergence that may have obscured the ancestral presence of a PRANC domain. A long history of rapid divergence is further supported by our findings that there is a large amount of variation in the number of ankyrin repeats, the number and position of introns, and chromosomal locations of the 15 DV ankyrin proteins (Additional file 1: Figure S1). We also confirmed that these proteins are distinct and not derived paralogs of canonical ankyrin domain proteins that are highly conserved in insects [18].

We decided to search for related proteins using the region 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 (*Nv*-CLANK-D and *Nv*-CLANK-E), which lack C-terminal sequence beyond the ankyrin domains. Since we knew that this region was predicted to contain a PRANC domain in four of our DV ankyrin proteins, we

hypothesized that the remaining C-termini maintained cryptic similarity to the ancestral PRANC domain.

Because the C-termini of our DV ankyrin domain proteins are less conserved, we used the more sensitive, iterative PSI-BLAST [19] approach to identify similar sequences in the NCBI non-redundant (nr) database. We used default parameters, including only using genes that were above the threshold in round 1 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 four rounds), as this was sufficient to identify the first non-insect and microbial sequences [20].

The taxa that appear in these queries are much more limited than what was found using the full DV ankyrin protein sequences as queries [21], indicating that complexity of aligning the constrained and repetitive ankyrin domains can give spurious signals of homology. The close relationship of *Nasonia* DV ankyrins (131 distinct sequences) and other orphan ankyrin domain-encoding genes in Chalcidoidea is accentuated, as multiple sequences from *Ceratosolen solmsi*, *Copidosoma floridanum*, and *Trichogramma pretosium* (13, 23, and 24 sequences, respectively) cluster more robustly with large numbers of *Nasonia* sequences when the aligning C-terminal regions are used for phylogenetic analysis, compared to analyses using the full-length proteins (Fig. 1a, b, compared to [21]). We also find sequences from species that also showed up strongly when using the full protein as a query [21], particularly from the ant *Pseudomyrmex gracilis* and the bee *Ceratina calcarata* (51 and 21 sequences, respectively). We also find a large number of hits from the Braconid wasp *Microplitis demolitor* and the whitefly *Bemisia tabaci* in all PSI-BLASTs (49 and 31 sequences, respectively). The sequences from these insects mostly cluster closely with others from the same species (Fig. 1a), indicating recent transfer and/or gene amplification events in these organisms.

Crucially, outside of these insect lineages, the only other significant hits were obtained in Rickettsial bacteria (primarily *Wolbachia* species and *Orientia tsutsugamushi*) and poxviruses (Fig. 1a, green and turquoise, respectively, [20]). While most of the bacterial sequences cluster together (Fig. 1a), some are embedded within clades composed primarily of chalcid wasps (Fig. 1b). While bootstrap support for this particular relationship is low, it is found consistently in multiple independent analyses [20]. Given that we propose that the horizontal transfer giving rise to the current distribution of lineage-restricted ankyrin domains in the chalcid wasps

occurred ancestrally over 100 million years ago, a significant loss of phylogenetic signal in this small protein domain is not surprising.

In contrast, a strongly supported relationship between 2 Wolbachia and all 51 Pseudomyrmex sequences (and 1 sequence from *Temnothorax curvispinosus*) was found (Fig. 1c), indicating that the horizontal transfer and gene amplification events were relatively recent and independent within this group.

Importantly, all of the hits aligned to the C-terminal region of proteins that also have ankyrin domains toward the N-termini of the proteins, indicating a conserved general structure with the proteins found now in insects.

Overall, these results strongly indicate that the ankyrin domain containing genes that were identified in the PSI-BLAST analyses have a complicated history. 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 events in lineages leading to the genera *Pseudomyrmex*, *Bemisia*, *Ceratina*, and *Microplitis*, and an ancient transfer in a common ancestor of the superfamily Chalcidoide (around 150 million years ago [22]). We propose to name this latter 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 *N. vitripennis* CLANKs (*Nv*-CLANK)-A through *Nv*-CLANK-O. The relationships between our CLANK nomenclature and gene identification numbers in different annotations are given in Additional file 1: Table S1.

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, 23, 24], 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 chalcid 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.

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 repeated and analyzed in more detail over a longer developmental

Rickettsia
CLANKS
Chalcid wasps
Bemisia

Pox viruses
Pseudomyrmex
Ceratina
Minor taxa

Fig. 1 (See legend on next page.)

time frame, and transcripts were grouped according to their expression patterns. Four CLANKS (*Nv-CLANK-B*, *Nv-CLANK-C*, *Nv-CLANK-D*, *Nv-CLANK-J*) have no patterned expression at any time in embryogenesis (Additional file 1: Figure S2) and will not be discussed further.

Laterally expressed CLANKs

Six of the 15 *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* (*Nv-CLANK-G*, *Nv-CLANK-H*, *Nv-CLANK-K*) show an overall expansion of expression, while three *Nv-CLANKs* (*Nv-CLANK-A*, *Nv-CLANK-E*, *Nv-CLANK-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. 2A1, A2). 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. 2A3) and then expands posteriorly creating a gradient with highest levels in that initial anterior domain (Fig. 2A4). Expansion is missing from both poles and also from the dorsal midline of the embryo (Fig. 2A5). During gastrulation, expression is restricted to a segmental pattern in the cells that give rise to the central nervous system (CNS) (Fig. 2A6).

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. 2B1, B2). Again, expression is absent at the AP poles and along the dorsal midline. However, while *Nv-CLANK-G* was expressed maternally, *Nv-CLANK-H* has no early expression (data not shown) and is ubiquitously expressed at very low levels during gastrulation instead of being localized to CNS precursors (Fig. 2B3).

Nv-CLANK-K is expressed maternally and ubiquitously at low levels (Fig. 2C1) before being expressed in an anterior-lateral domain with expression lacking at both the ventral and dorsal midline and strongest near the
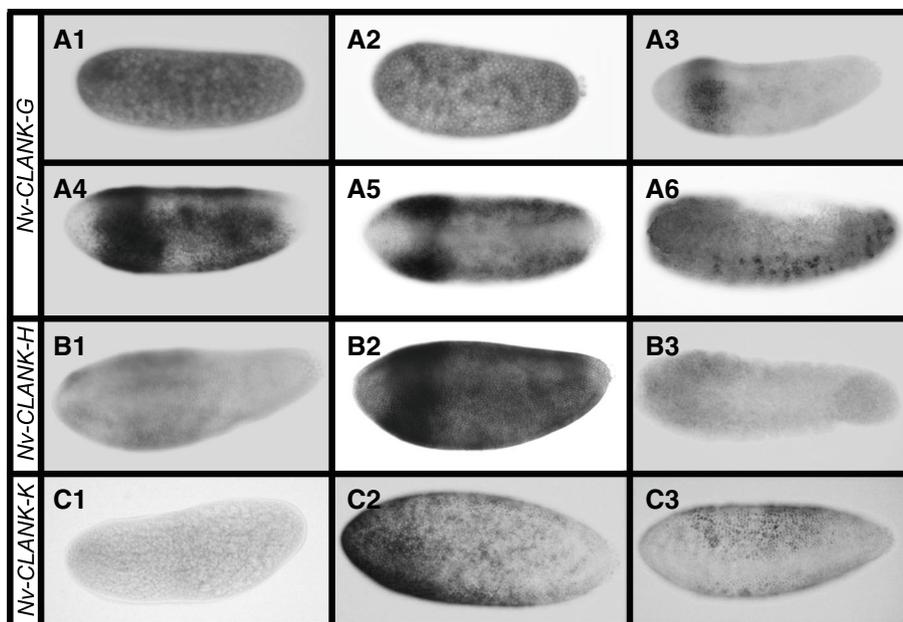
Fig. 2 Laterally expressed CLANKs with dynamic expansion. **A1–A6** Expression of *Nv-CLANK-G* from pre-blastoderm through gastrulation. **B1–B3** Expression of *Nv-CLANK-H* from blastoderm to gastrulation. **C1–C3** Expression of *Nv-CLANK-K* from pre-blastoderm through late blastoderm. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down (except A5, dorsal view)

anterior pole (Fig. 2C2). This domain then shifts and expands posteriorly into a more evenly expressed lateral domain, with inhibition of expression ventrally and at the poles (Fig. 2C3). During gastrulation expression is lost completely (data not shown).

Nv-CLANK-A is initially expressed ubiquitously at very low levels before becoming localized in a broad lateral domain in the early blastoderm (Fig. 3A1, A2). The lateral domain then shrinks into two discrete bands in the trunk of the embryo, before expanding first dorsoventrally and then anteroposteriorly into one lateral band (Fig. 3A3–A5). At times, expression is missing along the ventral and dorsal midlines; however, this is variable and dynamically changes as the blastoderm undergoes further division and cellularizes. Expression is lacking in the gastrulating embryo (Fig. 3A6).

Nv-CLANK-E expression is lacking or at very low levels both before and after the blastoderm stages (data not shown). Expression is in a broad band from nearly the middle of the embryo to just anterior of the posterior pole during the syncytial blastoderm (Fig. 3B1). During cellularization, this band retracts to a smaller, weaker, expression domain in the posterior of the embryo (Fig. 3B2, B3).

The expression pattern of *Nv-CLANK-I* in the mid to late blastoderm also resembles that of *Nv-CLANK-G* and *Nv-CLANK-H*. Again, there is a lateral expression domain with expression missing from the two poles and the dorsal midline (Fig. 3C1); however, the dynamics of this transcript differ. Instead of expression first appearing in the anterior region and then appearing progressively toward 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. 3C2) and then a posterior spot (Fig. 3C3). *Nv-CLANK-I* is not expressed maternally or in the early blastoderm (data not shown).

Dorsally expressed CLANKs

Three of the 15 transcripts are expressed dorsally during embryogenesis. *Nv-CLANK-N* and *Nv-CLANK-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. 4A1–A3). *Nv-CLANK-O* is expressed evenly along the dorsal midline from pole to pole, stably expressed throughout syncytial divisions and cellularization (Fig. 4B1–B2). While *Nv-CLANK-N* lacks expression during gastrulation, *Nv-CLANK-O* is expressed dorsally, surrounding the extraembryonic material (Fig. 4B3).

Nv-CLANK-F is initially expressed at a low ubiquitous level (Fig. 4C1) before gaining expression at the anterior and posterior poles (Fig. 4C2). Expression then expands along the dorsal midline, anterior to posteriorly, ultimately connecting the two poles (Fig. 4C2–C4). In
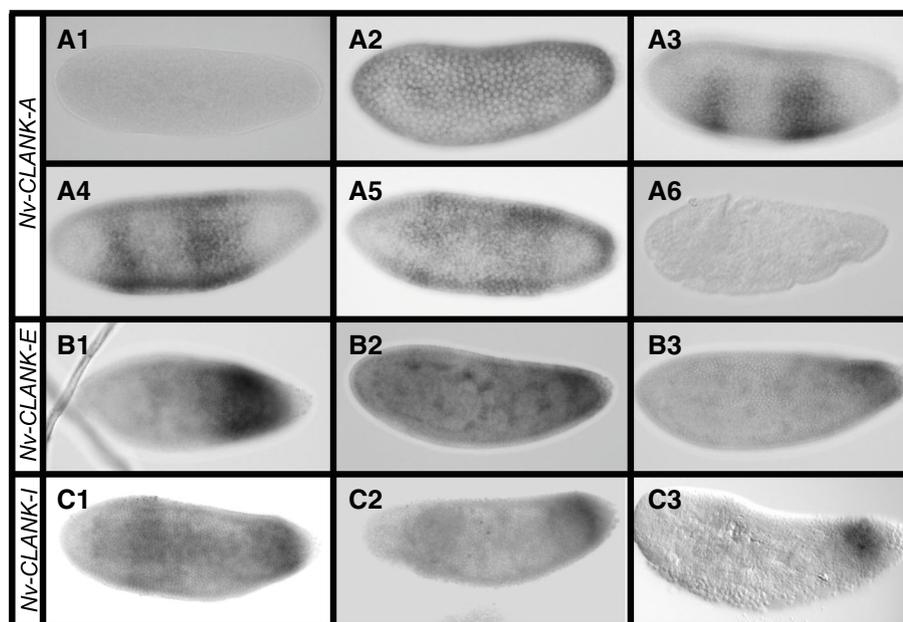
Fig. 3 Laterally expressed CLANKs with dynamic retraction. **A1–A6** Expression of *Nv-CLANK-A* from pre-blastoderm through gastrulation. **B1–B3** Expression of *Nv-CLANK-E* from mid to late blastoderm. **C1–C3** Expression of *Nv-CLANK-I* from blastoderm through gastrulation. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down
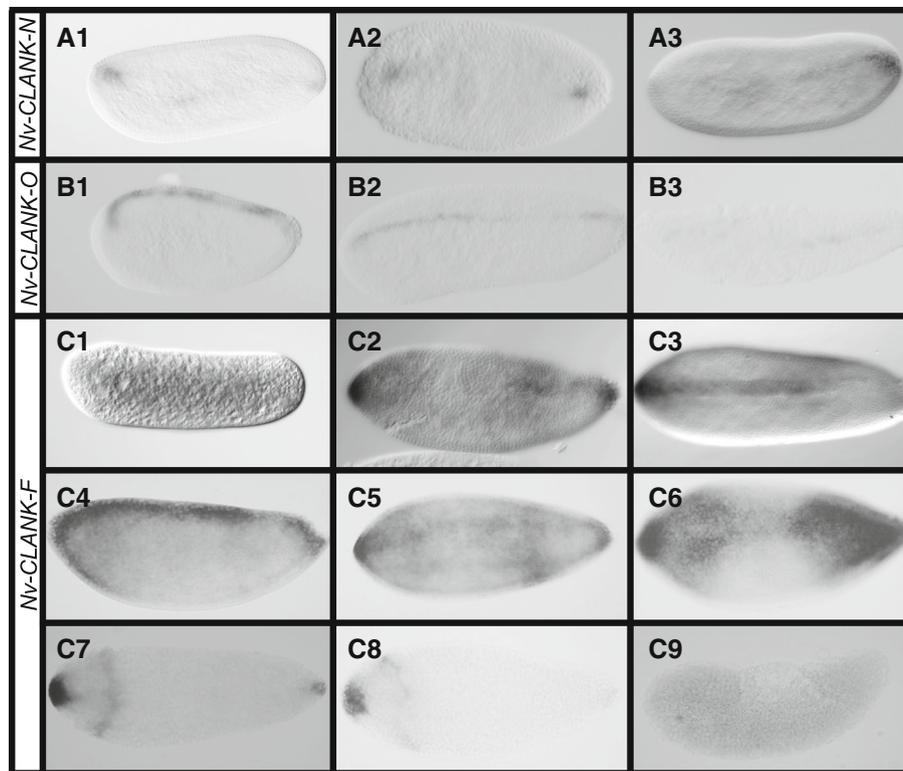
Fig. 4 Dorsally expressed CLANKs. **A1–A3** Expression of *Nv-CLANK-N* from early to late blastoderm. **B1–B3** Expression of *Nv-CLANK-O* from early blastoderm through gastrulation. **C1–C9** Expression of *Nv-CLANK-F* from pre-blastoderm through gastrulation. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down (except **C7** and **C8**, bird's eye dorsal view)

addition, there is dynamic ventral expansion that varies from embryo to embryo and from stage to stage. In some cases, the dorsal stripe expands past the anterior pole, onto the ventral half of the embryo, creating an anterior cap (Fig. 4C4). This expansion is accompanied by a perpendicular band, encircling the posterior end of the embryo's trunk (Fig. 4C5). In other cases, the expansion does not cross into the ventral half, but still expands to broad domains at the two poles, while remaining narrow in the mid-trunk region (Fig. 4C6). As the onset of gastrulation nears, this expansion, and the initial dorsal stripe, retracts, leaving a strong patch of expression at the anterior pole, a lighter and smaller domain at the posterior pole, and a weak stripe perpendicular to the dorsal midline in the anterior-dorsal region of the embryo (Fig. 4C7). While the anterior patch remains strong at the start of gastrulation, the other two expression domains weaken quickly (Fig. 4C8). Differential expression is eventually lost during gastrulation, and the whole embryo exhibits weak, ubiquitous expression (Fig. 4C9).

Ventrally expressed CLANK
Nv-CLANK-M is the only transcript in this family with ventral expression. It starts as a narrow stripe in the early blastoderm, much like *Nv-twist* (Fig. 5A1). Also,

like *Nv-twi*, it broadens later in development (Fig. 5A2). However, it never takes on a typical "slug" shape of the presumptive mesoderm and begins to disappear as gastrulation initiates. The pattern of disappearance roughly coincides with regions being covered by the lateral ectoderm (Fig. 5A3).

Ubiquitously and post gastrulation expressed CLANK
Nv-CLANK-L is strongly expressed both maternally (Fig. 5B1) and zygotically (Fig. 5B2–B3). Prior to gastrulation, expression is ubiquitous with the exception of there being no detectable expression within the budding pole cells (Fig. 5B2). During gastrulation, there is a moderate level of expression throughout the embryo; however, there are highly elevated levels of expression in the area that will become the head and in regions just anterior and posterior to the extraembryonic material (Fig. 5B3).

Reduction of CLANK transcripts results in significant increases in embryonic lethality
In order to understand the functional significance of this DV expression of the CLANKs in *Nasonia*, parental RNA interference (pRNAi), where double-stranded RNA is injected into female pupae and its effects are examined in the embryos she produces [25], was used to
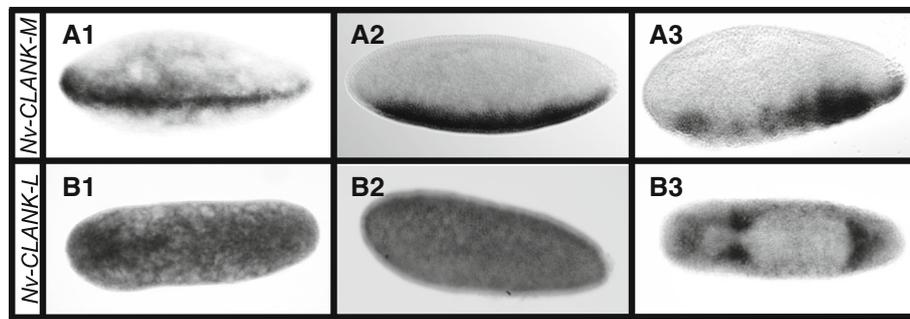
Fig. 5 Ventral and other CLANK expression patterns. **A1–A3** Expression of *Nv-CLANK-M* from early blastoderm through gastrulation. **B1–B3** Expression of *Nv-CLANK-L* from pre-blastoderm through gastrulation. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down (except B3, dorsal view)

knockdown each of the 11 genes with detectable DV expression patterns. We first analyzed the gross effects of knockdown on embryonic survival to hatching first instar larvae. Average embryonic lethality of the 11 knockdowns ranged from 0.87 to 12.19% of embryos plated (Fig. 6). In all cases, the frequency of lethality was larger than in control-injected pupae (0.65%). The difference in lethality was statistically significant ($p < 0.05$) for 6 of the 11 transcripts tested (Fig. 6).

In the course of these experiments focusing on embryonic development, we observed that the pupal injections also had significant and severe effects on successful pupal development for Nv-CLANKs. The effect was particularly strong in *Nv-CLANK-E*, *Nv-CLANK-K*, and *Nv-CLANK-L* where a quarter or less of injected pupae completed metamorphosis, compared to the 60% rate for mock-injected wasps (Additional file 1: Figure S3). This suggests that these transcripts may have additional functions in developmental or physiological processes of pupation.

Fig. 6 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 CLANK lethality to mock-injected lethality. Corresponding *P* values listed above graph (n.s. = non-significant)

CLANK transcript 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 1: Figure S4). *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 3 days post eclosion, we found large variability in the behavior of the dsRNAs (Additional file 1: Figure S4). 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 1: Figure S4).

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. In addition, while the CLANK transcripts are highly divergent at the nucleotide level, we cannot completely exclude potential effects of any dsRNA on non-target CLANKs.

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. 7A1, B1). This stripe stretches from the anterior to the posterior pole and is more or less equal in width and intensity throughout its domain. The blastoderm continues to divide the domain narrows (Fig. 7A2, B2) and eventually retracts from the posterior pole as the blastoderm cellularizes and gastrulation begins (Fig. 7A3, B3). During gastrulation, *Nv-zen* marks the serosa until it begins to migrate and encompass the embryo (Fig. 7A4, 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*, *Nv-CLANK-E*, *Nv-CLANK-F*, *Nv-CLANK-G*, *Nv-CLANK-H*, *Nv-CLANK-I*, *Nv-CLANK-K*, *Nv-CLANK-N*, *Nv-CLANK-O*) are knocked down, a number of changes in the expression of *Nv-zen* are observed (Fig. 7C1–F4, Additional file 1: Figure S5). 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. 7D3–4). 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. 7D1–2). For all of the knockdowns except for *Nv-CLANK-A* and *Nv-CLANK-K*, the increased frequency of reduced levels of *Nv-zen* expression was statistically significant ($p < 0.05$) (summarized in Fig. 8).
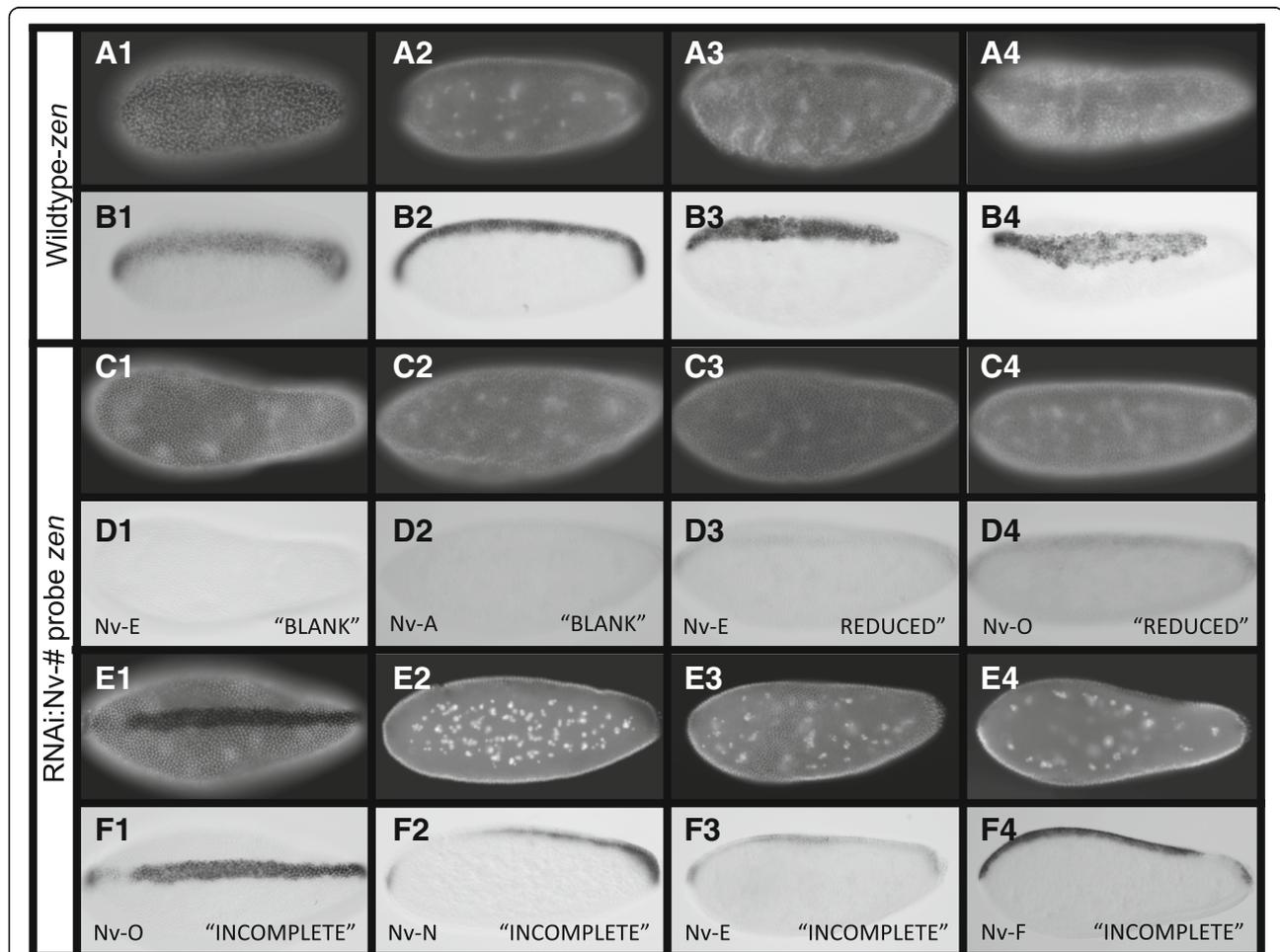
Fig. 7 Effects of reducing CLANKs on *Nv-zen* expression. **A1–B4** Expression of *Nv-zen* from early blastoderm through gastrulation in control embryos. **A1–A4** Control embryos stained with DAPI to approximate embryo age. **B1–B4** In situ hybridization of control embryos probing for *Nv-zen* expression. 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** and **E1–E4** Knockdown embryos stained with DAPI to approximate embryo age. **D1–D4** and **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** and **E1–E4**. All embryos are oriented with anterior to the left, posterior to the right, dorsal up, and ventral down

Fig. 8 Distribution of pRNAi phenotypes affecting *Nv-zen* expression for each knockdown condition. Percentage of knockdown embryos observed with wildtype *Nv-zen* expression, reduced levels of *Nv-zen* expression, an incomplete or partial dorsal stripe domain of *Nv-zen*, or lacking *Nv-zen* 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 the bar above graph, color corresponds to phenotype with significant difference). Schematic representation of each phenotype is shown below graph

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*, *Nv-CLANK-F*, *Nv-CLANK-N*, and *Nv-CLANK-O*. In mild cases, a small region adjacent to either the anterior or posterior pole is lacking expression (Fig. 7F1, F4), while the proximal pole and all distal regions appear unchanged. In more severe cases, larger regions, up to half the embryo (Fig. 7F2), or multiple regions throughout the embryo (Fig. 7F3) 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. 8).

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. 8).

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. 9A1–B1). As the blastoderm undergoes additional divisions, the stripe widens (Fig. 9A2–B2) before retracting at the anterior pole forming a pattern that resembles that of a slug (Fig. 9A3–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. 9B4) until the entire mesoderm is covered by neuroectoderm.

Ventral and laterally expressed CLANK transcripts (*Nv-CLANK-A*, *Nv-CLANK-E*, *Nv-CLANK-G*, *Nv-CLANK-H*, *Nv-CLANK-I*, *Nv-CLANK-K*, *Nv-CLANK-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. 9C1–F4, Additional file 1: Figure S6). 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. 9A1–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. 9C1–D2) 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 *Nv-CLANK-I* (Fig. 10).

The second phenotype observed in the early blastoderm is a delay in the expansion of *Nv-twi* (Fig. 9F1). The expansion of the *Nv-twi* domain is stereotyped and occurs between nuclear cycles 10 and 11 [13] (compare Fig. 9A1, A2, and E1). 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*, *Nv-CLANK-E*, *Nv-CLANK-G*, and *Nv-CLANK-M*. (Fig. 10).

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. 9B3). 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*, *Nv-CLANK-E*, *Nv-CLANK-G*, *Nv-CLANK-I*, and *Nv-CLANK-K* (significantly increased frequency in all but *Nv-CLANK-A*, Fig. 11), or to a much lower level than observed in wildtype embryos (*Nv-CLANK-G*, Fig. 9F2, significantly increased frequency, Fig. 11).

Nv-CLANK-A, *Nv-CLANK-E*, *Nv-CLANK-G*, *Nv-CLANK-H*, *Nv-CLANK-I*, or *Nv-CLANK-M* knockdowns
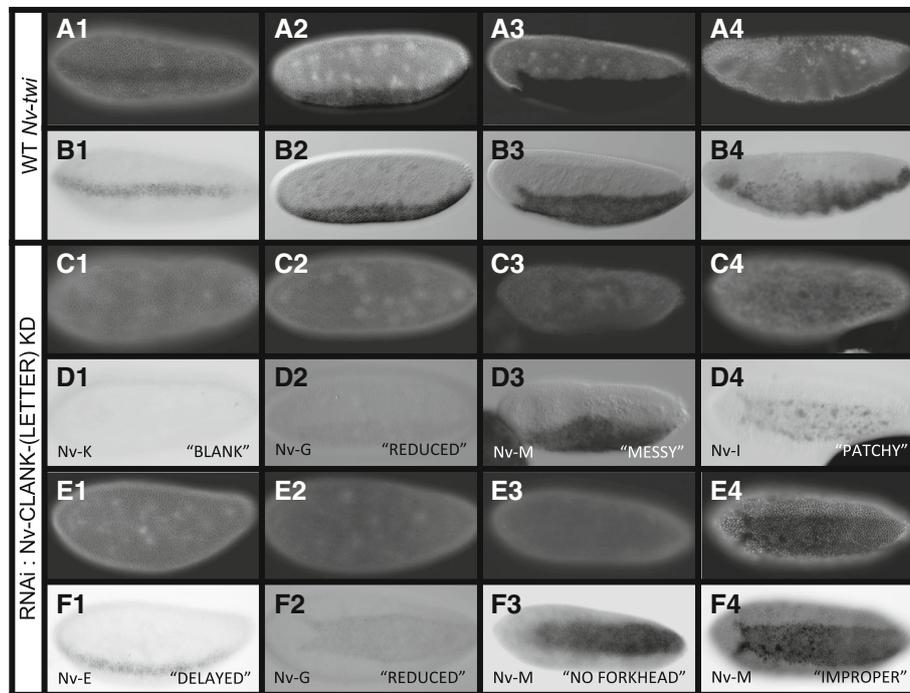
Fig. 9 Effects of reducing CLANKs on *Nv-twi* expression. **A1–B4** Wildtype expression of *Nv-twi* from early blastoderm through gastrulation. **A1–A4** Wildtype embryos stained with DAPI to approximate embryo age. **B1–B4** In situ hybridization of wildtype embryos probing for *Nv-twi* expression. Embryos **B1–B4** correspond to same embryos in **A1–A4**. **C1–F4** Altered expression of *Nv-twi* following pRNAi of one CLANK transcript (lower left corner) in early blastoderm through gastrulating embryos. **C1–C4** and **E1–E4** Knockdown embryos stained with DAPI to approximate embryo age. **D1–D4** and **F1–F4** In situ hybridization of knockdown embryos probing for *Nv-twi* expression ("phenotype" observed, lower right). Embryos correspond to the same embryos in **C1–C4** and **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)
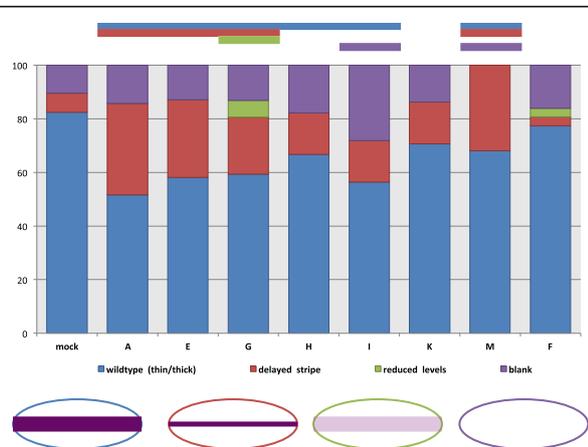
Fig. 10 Distribution of pRNAi phenotypes affecting early *Nv-twi* expression for each knockdown condition. Percentage of knockdown embryos observed with wildtype *Nv-twi* expression, a delay in the expansion of *Nv-twi* from a thin to thick ventral stripe, reduced levels of *Nv-twi* expression, or lacking *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 the bar above graph, color corresponds to phenotype with significant difference). Schematic representation of each phenotype is shown below graph

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*, *Nv-CLANK-E*, *Nv-CLANK-G*, *Nv-CLANK-H*, *Nv-CLANK-I*, and *Nv-CLANK-M* knockdown (Fig. 9D3). In rarer cases, *Nv-twi* domains where the anterior fork did not resolve are observed (only in *Nv-CLANK-G* and *Nv-CLANK-M*, Fig. 9F3). 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. 9D4). The size and number of these patches vary from embryo to embryo within and between knockdown 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. 11). The "messy" border and missing anterior fork phenotypes were never observed in wildtype embryos (Fig. 11).

The last time point where we looked for disruption is during gastrulation. Again, embryos were observed that

Fig. 11 Distribution of pRNAi phenotypes affecting 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 the 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 affecting 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 the bar above graph, color corresponds to phenotype with significant difference). Schematic representation of each phenotype is shown below graph

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. 12).

More interestingly, some knockdowns disrupted the morphogenetic movements of gastrulation. Normally mesoderm internalization proceeds from anterior to posterior in *Nasonia* [13] (Fig. 9B4). In rare instances, the mesoderm was observed internalizing posteriorly to anteriorly (Fig. 9F4) or in a random, disorganized manner (Additional file 1: Figure S6D, Q, Y, Z, EE) when *Nv-CLANK-A*, *Nv-CLANK-G*, *Nv-CLANK-K*, or *Nv-CLANK-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. 12).

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*, *Nv-CLANK-N*, *Nv-CLANK-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*, Figs. 10, 11, and 12; *Nv-CLANK-N* and *Nv-CLANK-O*, data not shown).

In summary, all of the 11 tested genes showed an increase in embryonic lethality compared to control. We could then show that markers of DV cell fates are disrupted spatially (for both *Nv-twi* and *Nv-zen*) and temporally (*Nv-twi* expansion) by knockdown of different CLANKs. This shows that these novel components of the DV GRN are functionally integrated and are important in producing a stable and reproducible patterning output.

The above results led us to wonder how long these genes have been a part of DV patterning in the wasp lineage to *Nasonia*. Are all of these genes unique recent additions to *Nasonia* DV GRN, or do some of them have a longer history in the wasp lineage?

Discovery of CLANKs in the wasp *M. digitata*
The second approach to understand the developmental and evolutionary significance of this gain of DV expression in *Nasonia* was to examine the function and expression of CLANKs in other species. This will help to understand how these genes have been functionally integrated into developmental processes.

As we described above, it appears that CLANKs are an ancestral and unique feature of the Superfamily Chalcidoidea. We have chosen to develop *M. digitata*, a representative of the Family Eulophidae (separated from *Nasonia* by about 90 million years of independent

evolution [26]), as a comparative model. *Melittobia* is attractive because it is easily reared in the lab on the same hosts as *Nasonia*, its mode of embryogenesis is rather similar to *Nasonia*, allowing for more straightforward 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 [27]. The sequences of the *potential Melittobia* CLANK homologs are presented in Additional file 1: Table S3 and were used to generate anti-sense probes to assess expression patterns (Fig. 13) and in phylogenetic analysis to assess their relationships among themselves and with *Nasonia* CLANKs (Fig. 14).

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 17 Md-CLANKs (*Md-CLANK-A*, *Md-CLANK-B*, *Md-CLANK-D*, *Md-CLANK-H*, *Md-CLANK-I*, *Md-CLANK-I2*, *Md-CLANK-J*, *Md-CLANK-K*, *Md-CLANK-L*, *Md-CLANK-N*), we identified as potential homologs of the DV Nv-CLANKS were not expressed differentially along the DV axis (Additional file 1: Figure S7).

In contrast, *Md-CLANK-C* again has dynamic expression in early and gastrulating embryos (Fig. 13A1–A6). It is absent in pre-blastoderm embryos (Fig. 13A1), then is initially expressed in three bands along the AP axis of the embryo (Fig. 13A2). 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. 13A3–A3'). This lateral domain then retracts into two discrete bands of expression (Fig. 13A4–A4'). The anterior most bands then disappears, leaving just one posterior band
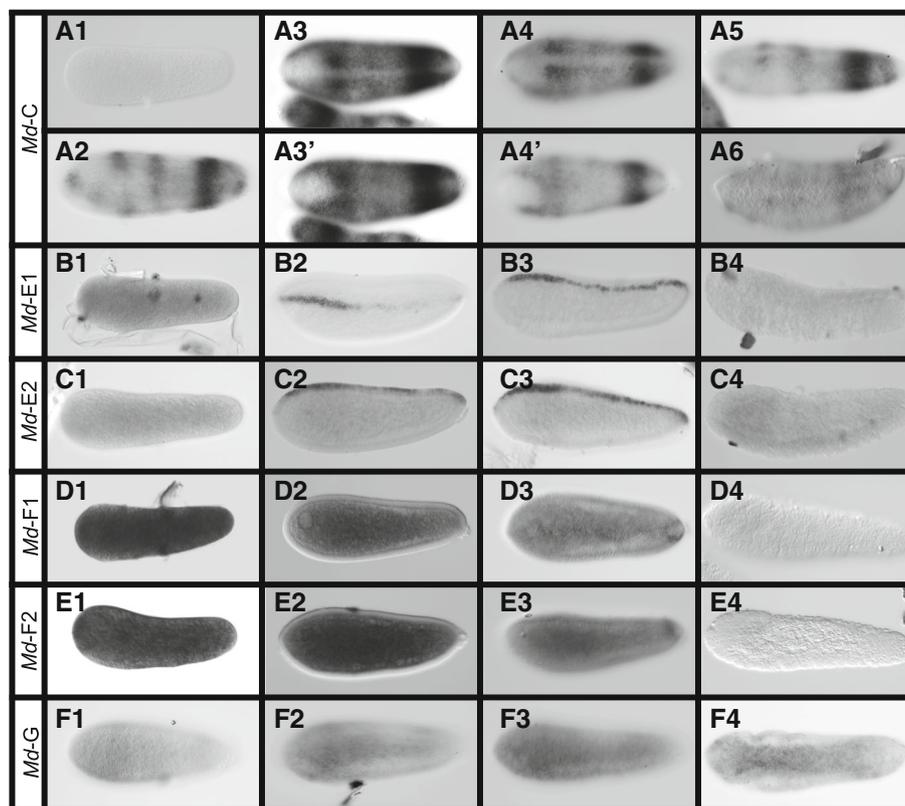
Fig. 13 *Melittobia* CLANK candidates with significant expression patterns. **A1–A6** Expression of *Md-CLANK-C* from pre-blastoderm through gastrulation. **A3**, **A4** Dorsal views. **A3'**, **A4'** Ventral views of the same embryo. **B1–B4** Expression of *Md-CLANK-E1* from pre-blastoderm through 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** 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. 14 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 [50]. Representative images of *Melittobia* orthologs with significant RNA localization and *Nasonia* ortholog with similar pattern (inset images). Colored box highlights pairing between wasp orthologs (violet, lime, teal, tangerine). Colored lines point to ortholog phylogenetic branch on tree (red = *Melittobia*, blue = *Nasonia*)

(Fig. 13A5). Throughout this retraction, the lack of staining at the dorsal midline and poles persists; however, there is staining at the ventral midline in the two bands (compare Fig. 13A3–A4'). During gastrulation, this posterior band of expression slowly fades until expression is lacking throughout the embryo (Fig. 13A6). This pattern is quite similar to previous patterns seen in *Nv-CLANK-G* and *Nv-CLANK-H* (Fig. 2A3–B2). These genes are in a phylogenetic cluster of *Nasonia* CLANKs that is a sister to a cluster of exclusively *Melittobia* CLANKs that contains *Md-CLANK-C* (Fig. 14).

Md-CLANK-E and *Md-CLANK-E2* also have low levels of ubiquitous expression in early embryos and no expression during gastrulation (Fig. 13B1, B4, C1, C4).

However, in blastoderm embryos, both are expressed in a stripe along the dorsal midline. The stripe is dynamic in expression levels and size along the entire AP axis for both *CLANKs*. Expression appears to originate at the anterior pole and fill in a discontinuous manner until the entire dorsal midline exhibits expression (Fig. 13B2, B3, C2, C3). Satisfyingly, these two genes cluster phylogenetically with *Nv-CLANK-O* (Fig. 14), which is expressed in an almost identical narrow dorsal stripe (Fig. 4B1, B2). This strongly indicates that the common ancestor of these genes was expressed dorsally and that this pattern has persisted for 90 million years.

Expression of *Md-CLANK-F* and *Md-CLANK-F2* is dynamic during blastoderm stages of development. Early pre-blastoderm and blastoderm-staged embryos have

light ubiquitous expression (Fig. 13D1, E1). Expression is then increased in the yolk of the syncytial blastoderm (Fig. 13D2, E2), reduced to low levels throughout, and then localized in a small patch of the dorso-posterior of the embryo (Fig. 13D3, E3) before quickly being lost again in cellularized blastoderm stage and gastrulating embryos (Fig. 13D4, E4). This pattern has no clear counterpart in the *Nasonia* genes we have examined.

In early blastoderm embryos, *Md-CLANK-G* expression appears to be weak and ubiquitous, with slightly higher expression on the ventral half of the embryo (Fig. 13F1). Expression increases in intensity forming a stripe along the ventral midline, widest in the anterior third of the embryo and narrowing in the posterior third (Fig. 13F2, F3). At the onset of gastrulation, expression is strongest and resembles the characteristic slug-shaped domain of *Nv-twist*, before staining is lost completely (Fig. 13F4). This pattern does not develop similarly to the ventrally expressed *Nv-CLANK-M* (Fig. 5A1–A3), and we do not consider it homologous.

The phylogenetic analysis revealed that there has likely been large-scale duplication and divergence (and/or gene conversion) in both wasp lineages (Fig. 14). Most of the *Nasonia* DV CLANKs cluster in two distinct clades on either side of the basal split of this protein tree. Similarly, most of the *Melittobia* proteins cluster together, or with the "off-target" *Nasonia* CLANKs that are not involved in DV patterning (Fig. 14). There are only a handful of cases that indicate clear orthology between a *Nasonia* DV CLANK and a *Melittobia* CLANK. *Nv-CLANK-O* clusters strongly with *Md-CLANK-E1* and *Md-CLANK-E2*. For the others, more complex evolutionary histories, involving ancestral genes that duplicated and diverged multiple times in both lineages after their separation must be considered. Thus, we propose that *Nv-CLANK-G*, Nv-CLANK-A, Nv-CLANK-H, Nv-CLANK-L, and Nv-CLANK-C derived the same common ancestral gene as *Md-CLANK-B*, *Md-CLANK-C*, *Md-CLANK-G*, *Md-CLANK-I1*, *Md-CLANK-I2*, and *Md-CLANK-K*. On the other hand, it is unclear how to relate *Nv-CLANK-F*, *Nv-CLANK-N*, *Nv-CLANK-M*, *Nv-CLANK-D*, *Nv-CLANK-I*, and *Nv-CLANK-J* to *Melittobia* counterparts.

Horizontally transferred and duplicating genes have been shown to be subject to complex processes of molecular evolution that make defining ancestry of genes particularly difficult [28], which might explain the difficulty in defining orthology in these relatively closely related species. In addition, potentially missing orthologs may have been misassembled in our transcriptome and thus not picked up by BLAST analysis. However, the genes we now have in hand are already quite informative as to the functional evolution of this gene family.

Discussion

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

Incorporation, duplication, and diversification of DV CLANKS

Ankyrin-repeat motifs (ANK) are important for protein-protein interactions and are commonly found in proteins across many species [24, 29]. Sequencing of the *Nasonia* genome uncovered that it contains the largest number of genes coding for ANK proteins of any insect [17]. Among this large number of ANK domain containing are orthologs to genes that are conserved features of insect genomes. However, the vast majority are orphan genes without clear orthologs in other insects that we have termed CLANKs in this manuscript. A clue to the microbial origin of the CLANKs was the discovery of PRANC domains at the C-termini of some of the proteins. The PRANC domain is found in *Wolbachia*, its bacteriophage, poxviruses, and various other bacteria, and its presence strongly indicated HGT from *Wolbachia* into the genome of an ancestor of *Nasonia*. Our results using PSI-BLAST to identify cryptic similarity of the C-termini of CLANKS that do not have annotated PRANC domains to C-termini of *Wolbachia* PRANC and ankyrin domain containing proteins further bolster this case.

Our observations indicate that the large number of CLANKs in *Nasonia* is the result mostly of duplication and divergence of genes present in the most recent common ancestor of the Chalcidoidea rather than repeated HGTs within the *Nasonia* lineage after the split among the families. This is based on our observation that the proteins are highly diverged from each other and from any presumed ancestor found in *Wolbachia*. In addition, the presence of introns in almost all of the sequences, their clear integration into the transcriptional regulation milieu of the *Nasonia*, and their dispersal throughout the genome strongly indicate that at least the 15 genes we detected as DV-regulated have a long history in wasp genomes and likely have arisen from duplication and divergence processes.

Apparently more recent HGT events may give clues about the origin and evolution of the CLANKs. For example, in the ant *Pseudomyrmex gracilis* [30] and the bee *Ceratina calcarata* [31], we find numerous ankyrin domain containing genes that cluster together in phylogenetic analyses and also cluster tightly with Wolbachia in these same analyses (Fig. 1a, c). We cannot exclude that multiple HGT events occurred in the lineage leading to *Nasonia* to give the full complement of CLANK genes in this wasp, and analyses to determine such facts should be an area of considerable effort in the future.

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

The chances of a gene horizontally transferred from a prokaryote to a eukaryote to be maintained in a genome is likely to be very low since not only must it gain the ability to be activated and processed by the eukaryotic transcriptional machinery, but it should also quickly gain a function in its new milieu. If these conditions are not met quickly, random mutations will accumulate and without selection, will eventually destroy protein-coding capacity of the transferred sequence (this is true whether a new gene arises by duplication, HGT, or de novo [32–36]).

Since ankyrin domains are protein binding domains [29, 37, 38] and direct protein-protein interaction is thought to be an important pre-cursor for proteins to gain new function [39, 40], ankyrin domain-encoding genes like the CLANKs may be predisposed to gain function in new environments. Furthermore, assuming the ancestral CLANK possessed a PRANC domain, it could have a ready-made interaction partner in the form of the Nf-κB homolog Dorsal, which has conserved roles in innate immunity and embryonic patterning throughout insects [41].

Why were CLANKS integrated into DV patterning?

One of the most surprising results of our previous analysis of DV patterning genes in *Nasonia* was the discovery of so many CLANKs with distinct and unique 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 [42]. Additionally, the PRANC domain has been described as being very similar to F-box domains [43], a domain that is known to induce ubiquitination of IkB, its degradation, and the activation of NF-kB [44]. These associations with NF-kB/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 [45], and CLANKs that are specifically upregulated in either male or female embryos, indicating a role in sex determination (JAL, personal observations). Further functional approaches will be undertaken to assess functional integration of CLANKs into these and additional GRNs.

What are the molecular roles of DV CLANKs?

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

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

At some level, it is not surprising that the CLANKs we find today have not maintained their hypothetical ancestral interaction partners, given the strong divergence of

these genes from each other at the amino acid level, and their 150 million years of evolution. Understanding the molecular interactions that mediate the function of the CLANKs will be a high priority in the coming years.

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

Our curiosity about the evolutionary history of the functional integration of CLANKs into the *Nasonia* DV patterning network led us to establish the wasp *Melittobia* as a satellite model organism. Our results indicate on the one hand that there is evidence that CLANKs have been integrated into the DV patterning GRN of some wasps for at least 90 million years (since the divergence of the Pteromalid and Eulophid Families) [26], based on the similarity of expression of two sets of CLANK homologs. On the other hand, we also revealed lineage-specific expression patterns for genes in both species, indicating that functional integration of CLANKs is an ongoing process. Sampling wasps in other Families in the Chalcidoidea will be necessary to pinpoint the origin of functional integration of DV CLANKs and to provide directionality to the changes (i.e., are some functional CLANKs being lost in some lineages, or is the pattern more due to independent gains?). It is likely that an unbiased approach to identify DV-regulated genes in *Melittobia* similar to the one we took in *Nasonia* [15], and broad application of this approach in a defined phylogenetic entity would allow for a high-resolution characterization of the evolutionary history of this gene family.

Conclusions

Our knockdowns of DV CLANKs showed that these genes significantly participate in ensuring successful embryogenesis and the formation of viable larvae. Further, we showed that a potential role of these genes is to ensure proper establishment of cell fates along the DV axis. However, neither the patterning nor the viability roles do the CLANKs appear to be absolutely essential. Rather, we propose that CLANKs act to constrain fluctuations in early development and that their loss can lead to variable fluctuations in patterning which in only rare cases lead to lethality. The source of the fluctuations could be environmental, genetic, or a combination of the two [6, 46, 47]. Our results show that even a very modest contribution to stability of development may lead novel GRN components to be maintained over significant evolutionary time periods. Thus, our results can be extrapolated to any potential novel components of GRNs, whether they originate from HGT, de novo genes, and co-option of existing genes into a new network.

Methods

BLASTs and sequence characterization

Our starting sequences were from the *Nasonia* annotation 2.0 [48]. Their NCBI counterparts were found by BLAST against the nt database [49]. These results provided the corresponding NCBI Reference Sequence Accession information for the *Nasonia* 2.0 annotations, which was used for downstream analyses of protein function.

Other similarity searches used blastp and searched the non-redundant protein database and default parameters, except cases where searches were limited to a single species.

Protein sequence alignment, phylogeny, and conserved domain analysis

Protein sequences corresponding to these 15 transcripts were then submitted to search for conservation at the amino acid level and to render a phylogenetic tree using "One Click Robust Phylogenetic Analysis" (http://phylogeny.lirmm.fr/) [50]. "One Click" parameters were as followed: Data & Settings (Gblocks not used), MUSCLE Alignment (-SeqType Protein), PhyML Phylogeny (Substitution model: WAG), and TreeDyn Tree Rendering (Reroot using mid-point rooting, Branch annotation: Branch support values). Finally, in order to gain family, domain, and repeat information for each transcript, we analyzed each protein using the Interpro Protein sequence analysis & classification software (InterProScan5) (https://www.ebi.ac.uk/interpro/) [51].

Phylogenetic analysis of the top 100 BLAST hits to each CLANK was performed at http://www.trex.u-qam.ca/ [52]. Alignments were performed using MUSCLE [53], alignments were manipulated for correct file type using AliView [54], relationships were inferred with RAxML [55], 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. Two iterations were performed for each gene, except CLANK-L, which required four 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 each other 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 (wildtype, cured of *Wolbachia*) pupa were injected with dsRNA (~ 1 μg/mL in water) designed against each of the transcripts with significant

DV expression as described in [25]. 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 h at 28 °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 h 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 h (at 28 °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 with in situ hybridization

In situ hybridization was performed using standard protocols [13, 15] on 0–24 h, 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 × 20 magnification on Zeiss widefield, compound epi-fluorescent microscope.

For knockdown experiments, in situ protocols were repeated on the 3–7 h (28 °C) knockdown embryos and on 3–7 h 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 to 7 h (28 °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. Twenty microliters per well PCR reactions were assembled using the PowerUp SYBR Green Master Mix (Applied Biosystems: A25742) (2X MM, 800 nM of each primer, cDNA, RFH2O). We performed the reactions in triplicate using the following parameters: (50 °C for 2′, 95 °C for 2′, 40 cycles of (95 °C for 15 s, 60 °C for 60 s, plate read, 72 °C for 60 s, plate read), 95 °C for 2′, gradient 60 °C➜95 °C (0.2 °C for 1 s).

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 3-day span. Standard deviation was used to calculate standard error, and T test ($P < 0.05$) was used to test for significance.

Melittobia sequences
Total mRNA (1 μg) libraries were created from various developmental time points (ovaries, pre-, early-, late blastoderm embryos, male-, female-yellow pupa) of the wasp *M. 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 BioProject database under accession number PRJNA429828 [56].

Melittobia ortholog discovery
We used a de novo assembled, unannotated embryonic *M. digitata* Transcriptome as a database for local tBLASTn (*Nv*-protein➜*Md*-mRNA) within the Geneious program (https://www.geneious.com/) [27] to search for orthologs to the *Nasonia CLANK* genes. A reciprocal BLAST was done on the top hits. If the reciprocal BLAST resulted in the *Nasonia* sequence that was input

into the query being returned as the top hit, the hit was considered a strong ortholog candidate. If it did not correctly BLAST back to the input sequence, the off-target *Nasonia* protein sequence returned was collected to be input in to downstream phylogenetic analysis. *Melitobia* transcript sequences were then translated into amino acid sequences via an online Translate tool (ExPASy, https://web.expasy.org/translate/). Longest ORF sequences were confirmed to align with tBLASTn hit sequences and then were collected for phylogenetic analysis.

Input *Nasonia*, off-target *Nasonia*, and all potential *Melittobia* ortholog protein sequences were then submitted to "One Click Robust Phylogenetic Analysis" (http://phylogeny.lirmm.fr/) [50], and trees were rendered relating each chalcid species orthologs to the *Nasonia* sequences as described earlier in our methods. Primers for cloning fragments of Melittobia sequences are given in Additional file 1: Table S4.

Melittobia in situ hybridization

Embryo collection, processing, and in situ protocol were developed and performed in a manner similar to *Nasonia* protocols with minor modifications. (Currently, in preparation for publication, specific details are available upon request.)

Additional file

Additional file 1: Table S1. Reference sequences and names corresponding to each novel ankyrin-repeat containing *Nasonia* transcripts. **Table S2.** CLANK ortholog candidate sequences in *Melittobia*. **Table S3.** *Nasonia* off-target hits from *Melittobia* CLANK ortholog BLAST. **Table S4.** *Melittobia* primer sequences for in situ hybridization probes. **Figure S1.** Phylogenetic analysis of *Nasonia* DV-regulated CLANK protein family. Figure S2. CLANKs lacking differential expression patterns. **Figure S3.** Distribution of pRNAi survival rate for each CLANK of interest. **Figure S4.** Relative embryonic expression of CLANK transcripts over time following pRNAi. **Figure S5.** Effects of reducing *CLANKs* on *Nv-zen* expression. **Figure S6.** Effects of reducing *CLANKs* on *Nv-twi* expression. **Figure S7.** *Melittobia* CLANK candidates lacking differential expression patterns. (PDF 11494 kb)

Abbreviations

ANK: Ankyrin-repeat motifs; CLANKs: Chalcidoidea Lineage specific ANKyrin domain-encoding genes; DV: Dorsoventral; GRN: Gene regulatory networks; HGT: Horizontal gene transfer; ISH: In situ hybridization; PRANC: Pox proteins Repeats of ANkyrin, C-terminal domain; pRNAi: Parental RNA interference; qPCR: Quantitative PCR; *twi*: *Twist*; *zen*: *Zerknüllt*

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

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Levine M, Davidson EH. Gene regulatory networks for development. Proc Natl Acad Sci U S A. 2005;102:4936–42.
2. Stathopoulos A, Levine M. Genomic regulatory networks and animal development. Dev Cell. 2005;9:449–62.
3. Moczek AP. Developmental capacitance, genetic accommodation, and adaptive evolution. Evol Dev. 2007;9:299–305.
4. Wilkins AS. Canalization: a molecular genetic perspective. Bioessays. 1997;19: 257–62.
5. Nijhout HF. 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;49:181–92.
6. von Dassow G, Meir E, Munro EM, Odell GM. The segment polarity network is a robust developmental module. Nature. 2000;406:188–92.
7. Moczek AP. On the origins of novelty in development and evolution. Bioessays. 2008;30:432–47.
8. Sommer RJ, Eizinger A, Lee KZ, Jungblut B, Bubeck A, Schlak I. 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;125:3865–73.
9. Abouheif E, Wray GA. Evolution of the gene network underlying wing polyphenism in ants. Science. 2002;297:249–52.
10. Wotton KR, Jimenez-Guri E, Crombach A, Janssens H, Alcaine-Colet A, Lemke S, Schmidt-Ott U, Jaeger J. Quantitative system drift compensates for altered maternal inputs to the gap gene network of the scuttle fly Megaselia abdita. Elife. 2015;4.
11. Wiegmann BM, Trautwein MD, Kim JW, Cassel BK, Bertone MA, Winterton SL, Yeates DK. Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. BMC Biol. 2009;7:34.
12. Davis GK, Patel NH. Short, long, and beyond: molecular and embryological approaches to insect segmentation. Annu Rev Entomol. 2002;47:669–99.
13. Buchta T, Ozuak O, Stappert D, Roth S, Lynch JA. Patterning the dorsal-ventral axis of the wasp Nasonia vitripennis. Dev Biol. 2013;381:189–202.
14. Ozuak O, Buchta T, Roth S, Lynch JA. Dorsoventral polarity of the Nasonia embryo primarily relies on a BMP gradient formed without input from Toll. Curr Biol. 2014;24:2393–8.
15. Pers D, Buchta T, Ozuak O, Wolff S, Pietsch JM, Memon MB, Roth S, Lynch JA. Global analysis of dorsoventral patterning in the wasp Nasonia reveals extensive incorporation of novelty in a regulatory network. BMC Biol. 2016;14.
16. Bordenstein SR, Bordenstein SR. Eukaryotic association module in phage WO genomes from Wolbachia. Nat Commun. 2016;7.

17. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Nasonia Genome Working G, Werren JH, Richards S, Desjardins CA, et al. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science. 2010;327:343–8.

18. Pers D, Lynch JA. Canonical ankyrin comparison [Internet]. In: figshare; 2018. https://doi.org/10.6084/m9.figshare.6983072.v1.

19. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.

20. Pers D, Lynch JA. C-terminal PSI BLAST tree [Internet]. In: figshare; 2018. https://doi.org/10.6084/m9.figshare.6981839.v1.

21. Pers D, Lynch JA. Full length CLANK BLAST and trees [Internet]. In: figshare; 2018. https://doi.org/10.1101/383968.

22. Machado CA, Jousselin E, Kjellberg F, Compton SG, Herre EA. Phylogenetic relationships, historical biogeography and character evolution of fig-pollinating wasps. Proc Biol Sci. 2001;268:685–94.

23. Penz T, Horn M, Schmitz-Esser S. The genome of the amoeba symbiont "Candidatus Amoebophilus asiaticus" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. Virulence. 2010;1:541–5.

24. Jernigan KK, Bordenstein SR. Ankyrin domains across the tree of life. Peerj. 2014;2.

25. Lynch JA, Desplan C. A method for parental RNA interference in the wasp Nasonia vitripennis. Nat Protoc. 2006;1:486–94.

26. Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, et al. Evolutionary history of the hymenoptera. Curr Biol. 2017;27:1013–8.

27. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.

28. Hao W, Palmer JD. HGT turbulence: confounding phylogenetic influence of duplicative horizontal transfer and differential gene conversion. Mob Genet Elem. 2011;1:256–61.

29. Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY. The ankyrin repeat as molecular architecture for protein recognition. Protein Sci. 2004;13:1435–48.

30. Rubin BER, Moreau CS. Comparative genomics reveals convergent rates of evolution in ant-plant mutualisms. Nat Commun. 2016;7.

31. Rehan SM, Glastad KM, Lawson SP, Hunt BG. The genome and methylome of a subsocial small carpenter bee, Ceratina calcarata. Genome Biol Evol. 2016;8:1401–10.

32. Chen S, Ni X, Krinsky BH, Zhang YE, Vibranovski MD, White KP, Long M. Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene. EMBO J. 2012;31:2798–809.

33. Zhang W, Landback P, Gschwend AR, Shen B, Long M. New genes drive the evolution of gene interaction networks in the human and mouse genomes. Genome Biol. 2015;16:202.

34. Lynch M, O'Hely M, Walsh B, Force A. The probability of preservation of a newly arisen gene duplicate. Genetics. 2001;159:1789–804.

35. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290:1151–5.

36. Assis R, Bachtrog D. Neofunctionalization of young duplicate genes in Drosophila. Proc Natl Acad Sci U S A. 2013;110:17409–14.

37. Sedgwick SG, Smerdon SJ. The ankyrin repeat: a diversity of interactions on a common structural framework. Trends Biochem Sci. 1999;24:311–6.

38. Bork P. Hundreds of ankyrin-like repeats in functionally diverse proteins—mobile modules that cross phyla horizontally. Proteins Struct Funct Genet. 1993;17:363–74.

39. Lynch M. The evolution of multimeric protein assemblages. Mol Biol Evol. 2012;29:1353–66.

40. Capra JA, Pollard KS, Singh M. Novel genes exhibit distinct patterns of function acquisition and network integration. Genome Biol. 2010;11.

41. Moussian B, Roth S. Dorsoventral axis formation in the Drosophila embryo—shaping and transducing a morphogen gradient. Curr Biol. 2005; 15:R887–99.

42. Chang SJHJ, 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;83:4140–52.

43. Mercer AA, Fleming SB, Ueda N. F-box-like domains are present in most poxvirus ankyrin repeat proteins. Virus Genes. 2005;31:127–33.

44. Spencer E, Jiang J, Chen ZJ. Signal-induced ubiquitination of IκBα by the F-box protein Slimb/β-TrCP. Genes Dev. 1999;13:284–94.

45. Quan H, Lynch JA. Transcriptomic and functional analysis of the oosome, a unique form of germplasm in the wasp Nasonia vitripennis. bioRxiv. 2018; 384032.

46. Kitano H. Biological robustness. Nat Rev Genet. 2004;5:826–37.

47. Gavin-Smyth J, Wang YC, Butler I, Ferguson EL. A genetic network conferring canalization to a bistable patterning system in Drosophila. Curr Biol. 2013;23:2296–302.

48. Rago A, Gilbert DG, Choi JH, Sackton TB, Wang X, Kelkar YD, Werren JH, Colbourne JK. OGS2: genome re-annotation of the jewel wasp Nasonia vitripennis. BMC Genomics. 2016;17.

49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

50. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008;36:W465–9.

51. Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.

52. Boc A, Diallo AB, Makarenkov V. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. Nucleic Acids Res. 2012;40: W573–9.

53. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

54. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30:3276–8.

55. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22:2688–90.

56. Pers D. Melittobia digitata embryonic transcriptome. In: BioProject; 2018. http://www.ncbi.nlm.nih.gov/bioproject/429828.