Article

Origin and elaboration of a major evolutionary transition in individuality

https://doi.org/10.1038/s41586-020-2653-6

Received: 2 October 2018

Accepted: 3 June 2020

Published online: 02 September 2020

Check for updates

Ab. Matteen Rafiqi[1,2,3], Arjuna Rajakumar[1,3] & Ehab Abouheif[1]✉

Obligate endosymbiosis, in which distantly related species integrate to form a single replicating individual, represents a major evolutionary transition in individuality[1–3]. Although such transitions are thought to increase biological complexity[1,2,4–6], the evolutionary and developmental steps that lead to integration remain poorly understood. Here we show that obligate endosymbiosis between the bacteria *Blochmannia* and the hyperdiverse ant tribe Camponotini[7–11] originated and also elaborated through radical alterations in embryonic development, as compared to other insects. The Hox genes *Abdominal A* (*abdA*) and *Ultrabithorax* (*Ubx*)—which, in arthropods, normally function to differentiate abdominal and thoracic segments after they form—were rewired to also regulate germline genes early in development. Consequently, the mRNAs and proteins of these Hox genes are expressed maternally and colocalize at a subcellular level with those of germline genes in the germplasm and three novel locations in the freshly laid egg. *Blochmannia* bacteria then selectively regulate these mRNAs and proteins to make each of these four locations functionally distinct, creating a system of coordinates in the embryo in which each location performs a different function to integrate *Blochmannia* into the Camponotini. Finally, we show that the capacity to localize mRNAs and proteins to new locations in the embryo evolved before obligate endosymbiosis and was subsequently co-opted by *Blochmannia* and Camponotini. This pre-existing molecular capacity converged with a pre-existing ecological mutualism[12,13] to facilitate both the horizontal transfer[10] and developmental integration of *Blochmannia* into Camponotini. Therefore, the convergence of pre-existing molecular capacities and ecological interactions—as well as the rewiring of highly conserved gene networks—may be a general feature that facilitates the origin and elaboration of major transitions in individuality.

The obligate endosymbiosis between the bacteria *Blochmannia* and ants of the Camponotini is thought to have contributed to the ecological and evolutionary success of these organisms[10,11,14–21]. Phylogenetic evidence suggests that the ancestor of *Blochmannia* was horizontally transferred from hemipteran bugs (a distantly related order of insects) to the most recent common ancestor of the Camponotini approximately 51 million years ago[10,16,22]. *Blochmannia* enhances nutrition by increasing amino acid synthesis, which can regulate the size distribution of worker ants[11,14,16,23]. Ants, in turn, provide *Blochmannia* with a protected cellular environment for proliferation and ensure the strict vertical transmission of these bacteria through the germline[17–19,24]. As a consequence, *Blochmannia* and Camponotini have co-evolved and their phylogenies are congruent[20,25,26].

Embryogenesis is radically altered

In ants, wasps and flies, the germplasm is a maternally inherited region of cytoplasm that is localized to the posterior pole of oocytes and freshly laid eggs, where it has a dual function in specifying the germline and the embryonic posterior[27–31]. The mRNAs and/or proteins of a group of highly conserved 'germline genes' are localized together in the germplasm[27] (Supplementary Table 1). To investigate whether the integration of *Blochmannia* into Camponotini influences the germplasm, we first determined the localization of mRNAs or proteins of germline genes in the freshly laid eggs of *Lasius niger*, an early-branching species that is in the same subfamily as the Camponotini (Formicinae) but that lacks *Blochmannia*. In *L. niger*, we found that *vasa* protein (Vas), *nanos* mRNA (*nos*), and *oskar* mRNA (*osk*) localize in a single germplasm at the posterior pole, similar to other ants, wasps and flies (Fig. 1a–c). We found that these germline genes in *Camponotus floridanus*, a species in Camponotini that has a germplasm surrounded by *Blochmannia*, also localize in a single germplasm at the posterior pole in oocytes (Extended Data Fig. 1a–i). Surprisingly, we discovered that, as the oocyte transitions to a freshly laid egg, the mRNAs or proteins of nine germline genes localize in four subcellular locations that we name 'zones': zone 1 (the ancestral position of the germplasm at the posterior pole at 100% length of the egg), zone 2 (located at about 80% egg length), zone 3 (located at about 60% egg length) and zone 4 (at the anterior

[1]Department of Biology, McGill University, Montreal, Quebec, Canada. [2]Present address: Beykoz Institute of Life Sciences and Biotechnology, Bezmialem Vakif University, Istanbul, Turkey. [3]These authors contributed equally: Ab. Matteen Rafiqi, Arjuna Rajakumar. ✉e-mail: ehab.abouheif@mcgill.ca

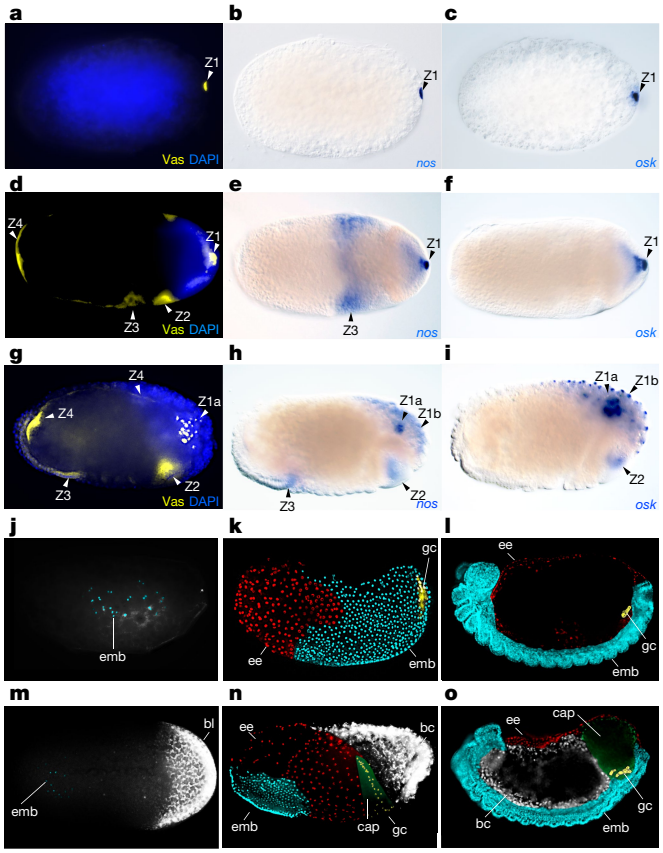


Fig. 1 | The evolution of four subcellular localization zones of germline genes that radically alter embryogenesis in *C. floridanus*. **a**–**c**, *L. niger* stage-1 freshly laid eggs, showing localization of Vas protein (**a**) in yellow, and *nos* mRNA (**b**) and *osk* mRNA (**c**) in blue. **d**–**f**, *C. floridanus* stage-1 freshly laid eggs, showing localization of Vas protein (**d**) in yellow, and *nos* mRNA (**e**) or *osk* mRNA (**f**) in blue. **g**–**i**, *C. floridanus* cellular blastoderm stage-6 embryos, showing expression of Vas protein (**g**) in yellow, and *nos* mRNA (**h**) and *osk* mRNA (**i**) in blue. **j**–**o**, Comparison of embryogenesis in *L. niger* (**j**–**l**) and *C. floridanus* (**m**–**o**). **j**, **m**, Freshly laid eggs (stage 1). **k**, **n**, Cellular blastoderm (stage 6). **l**, **o**, Segmentation (stage 12). False colouring highlights embryo (emb) in cyan, *Blochmannia* (bl) and bacteriocytes (bc) in white, germline (gc) in yellow, extraembryonic tissue (ee) in red and germline capsule (cap) in green. In **a**, **d** and **g**, blue indicates DAPI nuclear stain. Arrowheads indicate subcellular localization or expression zones of germline genes: zone (Z)1, zone 1a, zone 1b, zone 2, zone 3 and zone 4. Anterior is to the left; dorsal is to the top. In situ hybridization and immunohistochemistry experiments were repeated at least 4 (*L. niger*) or 8 times (*C. floridanus*) independently on $n \geq 5$ (*L. niger*) or $n \geq 30$ (*C. floridanus*) embryos per developmental stage.

pole at 0% egg length, extending along the dorsal side to the anterior boundary of *Blochmannia*) (Fig. 1d–f, Extended Data Fig. 1j–o). At a later stage, after the egg cellularizes and has initiated zygotic expression (the cellular blastoderm stage), the mRNAs or proteins of these nine genes persist in these four zones (Fig. 1g–i, Extended Data Fig. 1p–u). In both freshly laid and later-stage eggs, the localization and expression of mRNAs or proteins of germline genes is combinatorial—most are present in all four zones, but *nos* mRNA is only in two zones and *osk* mRNA is only in one zone (Fig. 1d–i, Extended Data Fig. 1j–u, Extended Data Table 1). Furthermore, the localization of these mRNAs or proteins is also dynamic: in later-stage eggs, the number of zones in which *nos* mRNA is present increases to three, and to two for *osk* mRNA—but the number of zones in which *smaug* (*smg*) mRNA is present decreases from four to three (Fig. 1d–i, Extended Data Fig. 1j–u, Extended Data Table 1). This combinatorial and dynamic localization shows that these four zones are not identical and suggests that they have distinct roles in integrating *Blochmannia* into *C. floridanus* during embryogenesis. Finally, because the freshly laid egg is a single host cell, the evolution of these four distinct zones is the result of changes in the subcellular localization of maternally inherited mRNAs and proteins.

We next asked how the evolution of the four zones of germline genes has affected embryogenesis in *C. floridanus*. We discovered that the eggs of *C. floridanus* are radically altered relative to those of other insects (Fig. 1j–o). Although insect embryos typically form at the posterior or throughout the entire egg[30,32], the embryo of *C. floridanus* forms in the anterior. At the posterior of the egg, *Blochmannia* become enveloped by specialized cells known as bacteriocytes and eventually migrate to the midgut to provide nutrition[11,14,16] (Fig. 1k, m–o, Extended Data Fig. 2a–c). Adjacent to bacteriocytes, the germ-cell precursors and a small population of *Blochmannia* are enveloped by a novel cell type we term the 'germline capsule', which—to our knowledge—has never previously been observed in insects (Fig. 1n, Extended Data Fig. 2d). The germline capsule then migrates posteriorly and attaches to the elongated embryo, where the germ cells and the small population of *Blochmannia* are transmitted to the next generation (Fig. 1o, Extended Data Fig. 2e, f). These results suggest that the four zones evolved to radically alter embryogenesis to integrate *Blochmannia* into *C. floridanus*.

We therefore investigated the role of each zone in this integration by tracking their fate in fixed embryos of known stages using Vas protein and *osk* mRNA (Extended Data Fig. 3). In freshly laid eggs, zone 1 initially appears as if it will form a posteriorly localized germplasm, as in other ants (Extended Data Fig. 3j, v). However, smaller germplasm foci begin budding off zone 1 and eventually give rise to two subzones: 'ancestral germplasm' at the posterior pole (which we term zone 1a) and germplasm foci at the centre of each bacteriocyte (zone 1b) (Extended Data Fig. 3a, b, j′, j″, m–p, v′, v″, w). At later stages, zone 1a and zone 1b migrate dorsally and are then no longer detectable (Extended Data Fig. 3c–h, p–u). By contrast, during the cellular blastoderm stage zone 2 is enveloped by the germline capsule and later migrates to connect to the embryonic posterior, where it gives rise to germ cells (Extended Data Fig. 3d–i, k, q–u, x). This shows that zone 2 is a novel germline of *C. floridanus*, and that the ancestral germplasm in zone 1 has lost its role in germline formation and acquired an alternative role within bacteriocytes. Next, zone 3 begins as a stripe in freshly laid eggs, and is later expressed throughout the germband, becoming enriched along the midline of the embryo; this suggests that zone 3 patterns the embryonic midline (Extended Data Fig. 3a–f). At this stage, zone 3 is also enriched at the posterior of the embryo, which suggests that it also specifies the embryonic posterior (Extended Data Fig. 3c–f). Finally, zone 4 is at the anterior pole in freshly laid eggs and then begins to extend dorsally, connecting to *Blochmannia* (Extended Data Fig. 3a, b). Later, zone 4 appears in the yolk membrane abutting the anteriormost cells of the embryo and extends all the way into the bacteriocytes (Extended Data Fig. 3c–e, l). Eventually, the yolk membrane forms the midgut that houses the bacteriocytes (Fig. 1o, Extended Data Fig. 2c). This suggests that zone 4 has a role in the migration of bacteriocytes to the midgut. Altogether, our data show that the four zones have distinct roles during the developmental integration of *Blochmannia* into *C. floridanus*—zone 1 and zone 4 have roles that are related to bacteriocytes; zone 2 is the functional germline; and zone 3 has a role in the embryonic midline and posterior. Zone 1 and zone 2 may have evolved to segregate *Blochmannia* into bacteriocytes for nutrition (zone 1) and into the germline capsule for vertical transmission (zone 2). Furthermore, zone 3 may have evolved to enhance the efficiency of this endosymbiosis by giving rise to an embryonic posterior in the anterior of the egg that is spatially separated from the *Blochmannia* populations in the posterior of the egg.

The Hox genes *abdA* and *Ubx* are rewired

In arthropods, the Hox genes *abdA* and *Ubx* function to morphologically differentiate the abdominal and thoracic segments after their

formation[33,34]. In hemipterans (from which the ancestor of *Blochmannia* colonized the Camponotini[10]), *abdA* and *Ubx* have an additional role in the development of bacteriocytes[35,36]. We therefore asked whether *abdA* and *Ubx* have a role in the integration of *Blochmannia* into *C. floridanus*. We discovered that—unlike in any other known insect— the mRNAs and proteins of *abdA* and *Ubx* in *C. floridanus* are localized in oocytes and freshly laid eggs, which shows that they are maternally inherited (Fig. 2a, d, Extended Data Fig. 4a–c). In freshly laid eggs, mRNAs and proteins of *abdA* localize in zone 1 and zone 3 and those of *Ubx* localize in all four zones (Fig. 2a, d, Extended Data Fig. 4c). At later stages, mRNAs and proteins of both genes are co-expressed with Vas protein in all four zones, and are also expressed in bacteriocytes (Fig. 2b, e, Extended Data Fig. 4d). Towards the end of embryogenesis, the conserved expression of *abdA* and *Ubx* appears in the third thoracic and abdominal segments (Fig. 2c, f, Extended Data Fig. 4e). Our results suggest that the mRNAs and proteins of *abdA* and *Ubx* interact with germline genes and have a role in all four zones.

To test this, we performed RNA interference (RNAi) on freshly laid eggs to knock down maternal and zygotic *abdA* and *Ubx* expression. We performed *abdA* RNAi at two concentrations, which produced a range of phenotypes in each zone as compared to a *YFP* RNAi control (Fig. 2g, h, j–n, Extended Data Fig. 4f–p). *abdA* RNAi at the lower concentration results in truncation of the embryonic posterior (zone 3) after the third abdominal segment (Fig. 2g, h). *abdA* RNAi at the higher concentration results in mild and severe phenotypes. Embryos with mild phenotypes develop into y-shaped embryos split along the ventral midline, and which—at later stages—truncate after the third abdominal segment (Fig. 2j–l, Extended Data Fig. 4f–k). Embryos with severe phenotypes develop into an embryonic stub or are undetectable (Extended Data Fig. 4l–p). Furthermore, *Ubx* RNAi results in the truncation of the embryonic posterior (zone 3) at the third thoracic segment (Fig. 2g, i). These results show that *abdA* and *Ubx* specify the embryonic posterior, and that *abdA* additionally functions in patterning the embryonic midline and forming the germband. Finally, *abdA* RNAi and *Ubx* RNAi also affect zone 1, zone 2 and zone 4: *Blochmannia* and bacteriocytes (zone 1 and zone 4) are eliminated (with *abdA* RNAi) or misplaced (with *Ubx* RNAi); the capsule (zone 2) develops external to the embryo (with *abdA* RNAi and with *Ubx* RNAi) or into an enlarged capsule (with *abdA* RNAi) (Fig. 2m–o, Extended Data Fig. 4o, p). Our RNAi data show that *abdA* and *Ubx* function in the four zones to integrate *Blochmannia* into *C. floridanus*.

We found that several germline genes were misexpressed after *abdA* RNAi, which suggests that *abdA* and *Ubx* are upstream of the germline genes (Extended Data Fig. 4f–n, p). To test this, we performed quantitative (q)PCR for nine germline genes on bacteriocytes (zone 1), the capsule (zone 2), and the germband and yolk sac together (zone 3 and zone 4) after dissecting them out of *YFP*-RNAi, *abdA*-RNAi, and *Ubx*-RNAi embryos (Extended Data Fig. 4q, r). We found that germline gene expression is downregulated in all three tissues after *abdA* RNAi and *Ubx* RNAi, and is significantly different from that in the *YFP*-RNAi control (Extended Data Fig. 4q, r). This shows that *abdA* and *Ubx* are rewired within the highly conserved segmentation hierarchy to regulate germline genes in the four zones.

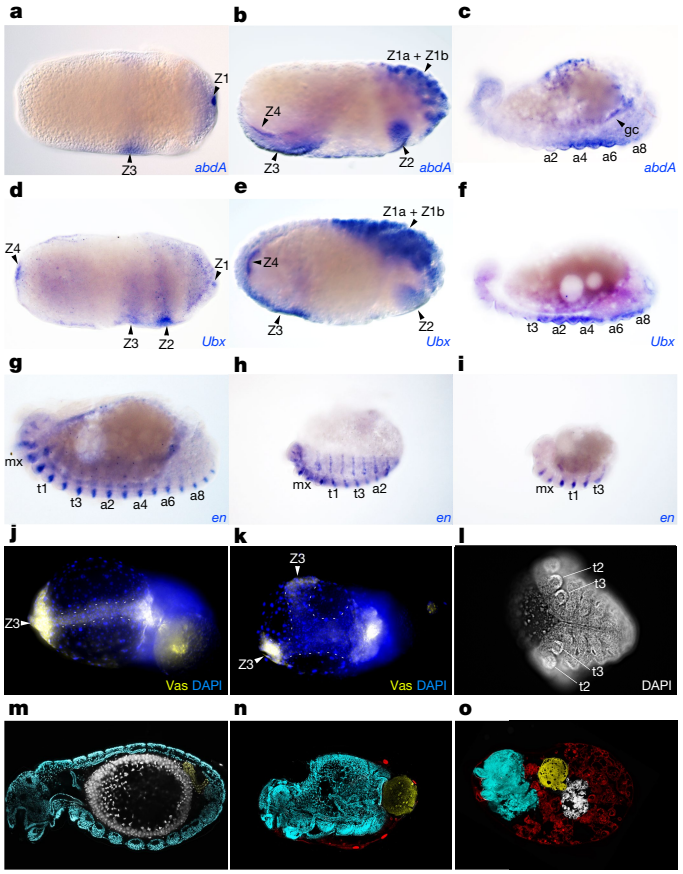


Fig. 2 | The Hox genes *abdA* and *Ubx* are rewired to regulate germline genes in *C. floridanus*. **a**–**f**, Wild-type *abdA* and *Ubx* mRNA staining in blue in freshly laid eggs (stage 1) (**a**, **d**), cellular blastoderm (stage 6) (**b**, **e**) and segmented (stage 12) (**c**, **f**) embryos. **g**–**i**, Stage-12 embryos with *engrailed* (*en*) staining in blue, showing the control *YFP* RNAi phenotype ($n = 70, 100\%$) (**g**), the low-concentration *abdA* RNAi phenotype ($n = 35$ out of 63, 56%) (**h**) and the high-concentration *Ubx* RNAi phenotype ($n = 113$ out of 122, 93%) (**i**). **j**, **k**, Stage-8 eggs with Vas protein in yellow, DAPI in blue and the embryo marked by dotted lines, showing the control *YFP* RNAi phenotype ($n = 45, 100\%$) (**j**) and the high-concentration *abdA* RNAi phenotype ($n = 21$ out of 61, 34%) (**k**). **l**, Stage-12 eggs, showing the high-concentration *abdA* RNAi phenotype with DAPI in white ($n = 22$ out of 31, 71%). **m**–**o**, Stage-17 eggs false-coloured to show embryo (cyan), serosa (red), *Blochmannia* and bacteriocytes (white), and germline capsule (yellow), showing the control *YFP* RNAi phenotype ($n = 70$, 100%) (**m**), the low-concentration *abdA* RNAi phenotype ($n = 35$ out of 63, 56%) (**n**) and the high-concentration *Ubx* RNAi phenotype ($n = 113$ out of 122, 93%) (**o**). Segments are marked as: maxillary (mx), thoracic segments (t)1–3 and abdominal segments (a)1–8. Zones and subzones are indicated with arrows. Anterior is to the left, dorsal is to the top; except for **j**–**l**, in which ventral is towards the reader. In situ hybridization and immunohistochemistry experiments (**a**–**f**) were repeated at least 8 times independently on $n \geq 30$ embryos per developmental stage.

Blochmannia regulate Hox and germline genes

Blochmannia makes up 97.2% of the total DNA content in freshly laid eggs (Extended Data Fig. 2g–k). We therefore tested whether *Blochmannia* influences the four zones. To do this, we treated *C. floridanus* colonies with rifampicin, an antibiotic that eliminates *Blochmannia*[11,14,15,19] (Fig. 3a, b, Extended Data Fig. 5a, c). We discovered that the formation of all four zones in freshly laid eggs is unaffected (Fig. 3c, Extended Data Fig. 6a, g, j, m, p). However, *abdA* mRNA and *tudor* protein (Tud) are lost from zone 1 and the ancestral germplasm becomes more tightly localized at the posterior pole, resembling the germplasm in other ants, wasps and flies[29–31] (Fig. 3a, f, Extended Data Fig. 6a, d, g, j, m, p). Therefore, in freshly laid eggs, all four zones are established by *C. floridanus*, while *Blochmannia* selectively regulates mRNAs and proteins to modify zone 1. At the cellular blastoderm stage, we observed a range of phenotypes that we categorized into two classes ('severe' and 'mild') that occur in equal proportion. Severe phenotypes are nonviable because they have no developing germband, show a range of morphological defects, and in all four zones, the mRNAs or proteins of *abdA*, *Ubx* and the germline genes are either absent or mislocalized (Fig. 3e, h, Extended Data Fig. 6c, f, i, l, o, r). By contrast, mild phenotypes are viable and show no morphological defects, and the mRNAs

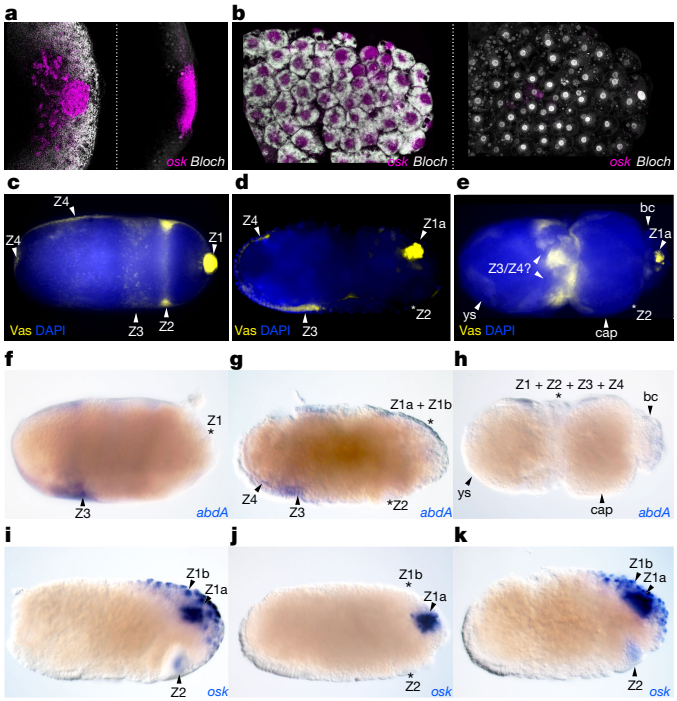


Fig. 3 | *Blochmannia* bacteria maintain and selectively regulate mRNA and proteins of maternal Hox and germline genes. a, Stage-1 freshly laid eggs stained with *osk* mRNA in magenta and DAPI, which marks *Blochmannia* (*Bloch*) in white, showing wild-type phenotype ($n \geq 30$) (left) and rifampicin-treated phenotype ($n \geq 15$) (right). **b**, Bacteriocytes from stage-6 embryos stained with *osk* mRNA in magenta and DAPI in white, showing wild-type phenotype ($n \geq 30$) (left) and rifampicin-treated phenotype ($n \geq 15$) (right). **c**–**e**, Embryos from rifampicin-treated colonies, showing Vas protein in yellow and DAPI in blue in a stage-1 freshly laid egg ($n = 29$) (**c**), a stage-6 embryo showing a mild phenotype ($n = 55$) (**d**) and a stage-6 embryo showing a severe phenotype ($n = 58$) (**e**). **f**–**h**, Embryos from rifampicin-treated colonies, showing *abdA* mRNA in blue in a stage-1 freshly laid egg ($n \geq 15$) (**f**), a stage-6 embryo showing a mild phenotype ($n \geq 15$) (**g**) and a stage-6 embryo showing a severe phenotype ($n = 6$) (**h**). **i**–**k**, Comparison of morphology and *osk* mRNA expression (blue) between stage-6 wild-type embryos (**i**), stage-6 embryos with a mild phenotype ($n = 39$) (**j**) and stage-6 embryos transplanted with *Blochmannia* ($n = 35$) (**k**), which were collected from the same *Blochmannia*-free rifampicin-treated colony as in **j**. One hundred per cent of the transplanted embryos develop into phenotypes similar to wild type ($n = 35$ of 35; compare **i** with **k**), and 0% ($n = 0$ of 35) develop into mild or severe phenotypes, one-tailed Fisher's exact test (degrees of freedom = 1, $P = 0.00002$). Asterisks indicate absence of localization or expression in zones or subzones. Arrowheads indicate zones or subzones. bc, bacteriocytes; cap, giant capsule; ys, yolk sac. Question marks indicate presumptive zones. Anterior is to the left, dorsal is to the top. In situ hybridization and immunohistochemistry experiments were repeated at least eight times independently.

or proteins of *abdA*, *Ubx* and the germline genes are selectively lost in each of the four zones: *abdA* and Tud in zone 1a; *abdA*, *Ubx*, *osk*, *nos* and Tud in zone 1b; *abdA*, *Ubx*, *osk*, *nos*, Vas and *aubergine* protein (Aub) in zone 2; and *staufen* mRNA (*stau*) in zone 3 and zone 4 (Fig. 3d, g, Extended Data Fig. 6b, e, h, k, n, q, Extended Data Table 1). At later stages, mild-phenotype embryos develop normally but often show defects in the gonads (Extended Data Fig. 5j, l, m, o–q). These results show that *Blochmannia* bacteria maintain mRNAs or proteins within each zone and selectively regulate them to make each zone functionally distinct.

To rule out the possibility that these changes are the unspecific effect of antibiotic treatment, we transplanted *Blochmannia* from wild-type eggs into *Blochmannia*-free eggs from a rifampicin-treated colony. One hundred per cent of the transplanted embryos developed into embryos similar to those of wild-type eggs, with *osk* mRNA restored to both zone 1b and zone 2; by contrast, un-transplanted control embryos from the same rifampicin-treated colony developed mild and severe phenotypes (Fig. 3i–k). Furthermore, we treated *C. floridanus* with ampicillin and a *Blochmannia*-free species (*L. niger*) with rifampicin. These embryos developed into embryos that are similar to those of the wild type in each case (Extended Data Fig. 5). Therefore, our results show that, after all four zones are established by *C. floridanus*, *Blochmannia* bacteria selectively regulate the mRNAs and proteins of *abdA*, *Ubx* and the germline genes to make the four zones functionally distinct and to maintain these zones. Because *abdA* and *Ubx* are upstream of the germline genes, *Blochmannia* selectively regulate germline genes through *abdA* and *Ubx*.

The origin and elaboration of integration

On the basis of our results, we predicted that the radical alterations we observed in *C. floridanus* embryos evolved in the most recent common ancestor of the Camponotini during the origin of the obligate endosymbiosis with *Blochmannia*. To test this prediction, we used RevBayes[37] to reconstruct the ancestral states of five developmental characters within the subfamily Formicinae (Fig. 4a, Extended Data Figs. 7, 8) to uncover the origin and elaboration of the developmental integration of *Blochmannia* into the Camponotini. In the ancestors of basally branching lineages, our reconstruction infers an embryo located in the posterior of the egg with a single germplasm (Fig. 4a (nodes 0–9 and 12–16), Extended Data Fig. 7 w–y, za, zb, zc, zd, ze). Notably, at node 10, node 11 and in *Brachymyrmex patagonicus*, the embryo shifted its location to the anterior but retained a single germplasm at the posterior of the egg (Fig. 4a, Extended Data Fig. 7t–v, z). Furthermore, the most recent common ancestor of the four closest sister tribes of the Camponotini evolved a novel subcellular localization zone for Vas and the maternal AbdA and Ubx proteins (Fig. 4a (nodes 12–16), Extended Data Figs. 7o–s, zf, zg, 8l–p). We infer that this zone is homologous to zone 3, because it is in a position similar to zone 3 in *C. floridanus* embryos and lacks *osk* mRNA (which exclusively marks zone 1 and zone 2 in *C. floridanus*) (Extended Data Fig. 7zh, zi, zj, zk). Finally, in addition to the Camponotini, different obligate endosymbionts evolved independently in the most recent common ancestor of the Formicini and Plagiolepidini[10,25,38] (Fig. 4a, Extended Data Fig. 7zf′, zg′). Therefore, the ability to shift the embryo to the anterior and the capacity to localize mRNAs and proteins to novel zones evolved before the three known obligate endosymbioses in ants at node 12 (Fig. 4a).

At the origin of the obligate endosymbiosis between *Blochmannia* and Camponotini, our reconstruction infers the evolution of three innovations: maternal AbdA and Ubx now localize to the ancestral germplasm (zone 1); zone 4 appears at the anterior pole and localizes Vas and maternal AbdA and Ubx; and the embryo shifts to the anterior of the egg, forming a novel embryonic posterior within zone 3 (Fig. 4a (node 17), Extended Data Figs. 7f–n, 8e–k). This integration was later elaborated within the derived genus *Camponotus* with two additional innovations: a germline in zone 2 that localizes Vas and maternal AbdA and Ubx and its surrounding germline capsule (Fig. 4a, Extended Data Figs. 7a–e, 8a–d). Our reconstruction uncovers the innovations that evolved before, during and after the origin of the obligate endosymbiosis between *Blochmannia* and Camponotini.

Discussion

Here we provide evidence for the following pathway for the origin and elaboration of developmental integration between *Blochmannia* and Camponotini (Fig. 4b). In step 1 (pre-existing capacity), a novel zone (zone 3) evolved to have a role in embryonic patterning, before the origin of this developmental integration. This led to a pre-existing capacity to localize mRNAs and proteins to novel subcellular locations,

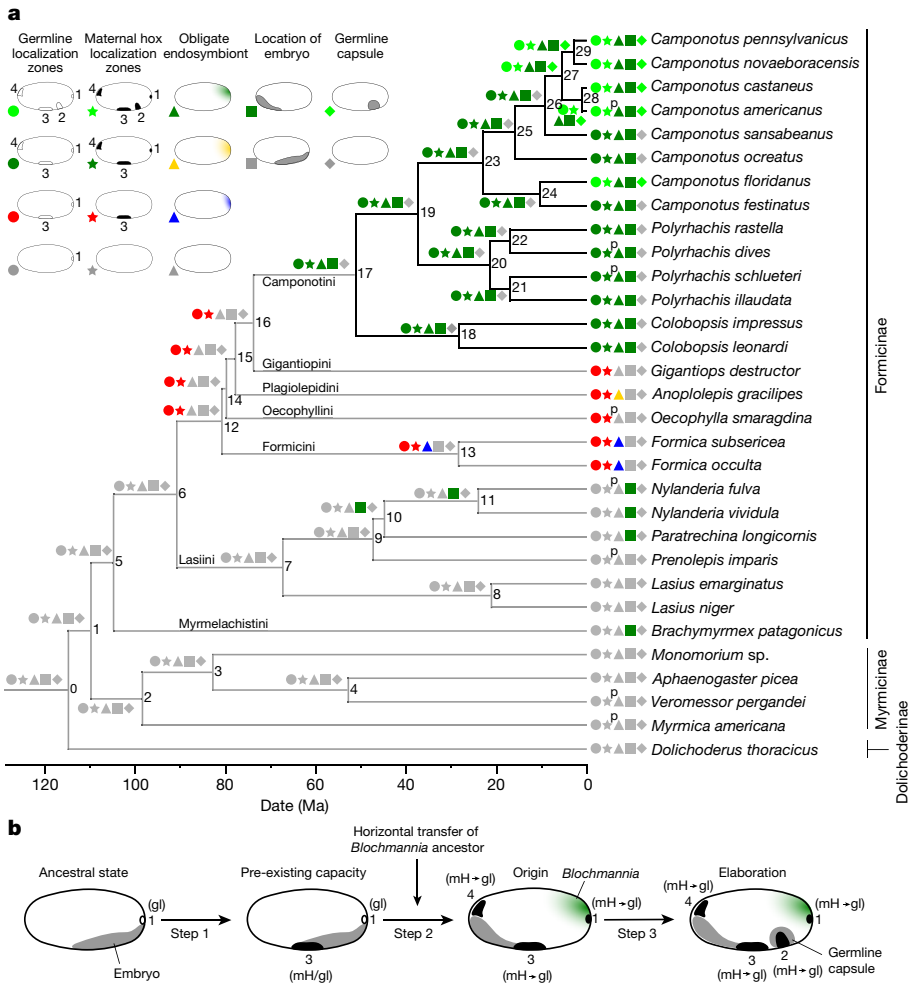


Fig. 4 | Origin and elaboration of developmental integration of *Blochmannia* into Camponotini. a, Phylogenetic tree of species within the ant subfamily Formicinae and outgroups; scale indicates millions of years ago (Ma). Nodes are numbered from 0 to 29. Species and subfamily names are indicated to the right of the tree, and names of tribes are indicated nearest to their node of origin. Black branches indicate lineages within the Camponotini with an obligate endosymbiosis with *Blochmannia*. Zones of mRNAs and proteins of germline genes are indicated by circles, and those of the maternal Hox genes *abdA* and *Ubx* are indicated by stars. Obligate endosymbionts in the posterior that are from different taxonomic lineages are indicated by triangles, the position of the embryo is indicated by a square and the presence of a germline capsule is indicated by a diamond. The different states of a developmental character are indicated by the different colours of the shaped icons. Character states predicted by RevBayes are marked with a 'p' superscript. Posterior probabilities for each node are listed in Extended Data Table 2. **b**, Proposed steps in the origin and elaboration of developmental integration between *Blochmannia* and the Camponotini. The number of localization zones (1 to 4) of mRNAs and proteins of germline genes (gl) and of the maternal Hox genes *abdA* and *Ubx* (mH) are indicated in black. Embryo and germline capsule are indicated in grey; *Blochmannia* is indicated in light green. Solidus (mH/gl) indicates that the maternal Hox genes are co-expressed with the germline genes, and arrows indicate that the maternal Hox genes are upstream of the germline genes.

which was subsequently co-opted to facilitate the integration of *Blochmannia* into the Camponotini. Furthermore, colocalization of *abdA*, *Ubx* and the germline genes in zone 3 facilitated the rewiring of *abdA* and *Ubx* to regulate the germline genes either before or at the origin of integration. In step 2 (origin), *Blochmannia* gained the ability to selectively regulate germline genes through *abdA* and *Ubx* in each zone, which led to the evolution of three functionally distinct zones: the ancestral germline (zone 1); the embryonic midline and posterior (zone 3), which allowed the shift of the embryo to the anterior; and zone 4, which guides bacteriocytes to the midgut. In step 3 (elaboration), some derived *Camponotus* species evolved a novel germline (zone 2) surrounded by a germline capsule, which freed the ancestral germline (zone 1) to have an alternative role within bacteriocytes. Therefore, the origin and elaboration of this major transition in individuality occurred through the stepwise addition of zones from 1 to 2 to 3 to 4 (Fig. 4b). This stepwise addition of zones evolved through tinkering with subcellular localization to produce distinct modules that divide labour within a single cell through the combinatorial localization of the same genes. Finally, the ecological mutualism between hemipteran bugs and the Camponotini[10,12,13] is thought to have facilitated the horizontal transfer of *Blochmannia* into the Camponotini, which suggests that ecological circumstances and pre-existing developmental capacities must converge to produce favourable conditions for major evolutionary transitions to obligate endosymbiosis. We therefore propose that other major transitions in individuality may originate and also elaborate through the rewiring of highly conserved gene regulatory networks, as well as by exploiting pre-existing molecular or developmental capacities and ecological interactions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2653-6.

1. Maynard-Smith, J. & Szathmary, E. *The Major Transitions in Evolution* (Oxford Univ. Press, 1997).
2. West, S. A., Fisher, R. M., Gardner, A. & Kiers, E. T. Major evolutionary transitions in individuality. *Proc. Natl Acad. Sci. USA* **112**, 10112–10119 (2015).
3. Sachs, J. L., Skophammer, R. G. & Regus, J. U. Evolutionary transitions in bacterial symbiosis. *Proc. Natl Acad. Sci. USA* **108** (Suppl 2), 10800–10807 (2011).
4. Boomsma, J. J. & Gawne, R. Superorganismality and caste differentiation as points of no return: how the major evolutionary transitions were lost in translation. *Biol. Rev. Camb. Philos. Soc.* **93**, 28–54 (2018).
5. Moran, N. A. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc. Natl Acad. Sci. USA* **104** (Suppl 1), 8627–8633 (2007).
6. Ruiz-Trillo, I. & Nedelcu, A. M. *Evolutionary Transitions to Multicellular Life: Principles and Mechanisms* Vol. 2 (Springer, 2015).
7. Blochmann, F. Über das Vorkommen bakterienähnlicher Gebilde in den Geweben und Eiern verschiedener Insekten. *Zbl. Bakteriol.* **11**, 234–240 (1892).
8. Buchner, P. *Endosymbiosis of Animals with Plant Microorganisms* (Interscience, 1965).
9. Tanquary, M. C. *Biological and Embryological Studies on Formicidae*. PhD thesis, Univ. of Illinois (1912).
10. Wernegreen, J. J., Kauppinen, S. N., Brady, S. G. & Ward, P. S. One nutritional symbiosis begat another: phylogenetic evidence that the ant tribe Camponotini acquired *Blochmannia* by tending sap-feeding insects. *BMC Evol. Biol.* **9**, 292 (2009).
11. Zientz, E., Beyaert, I., Gross, R. & Feldhaar, H. Relevance of the endosymbiosis of *Blochmannia floridanus* and carpenter ants at different stages of the life cycle of the host. *Appl. Environ. Microbiol.* **72**, 6027–6033 (2006).
12. Aranda-Rickert, A., Fracchia, S., Yela, N. & Marazzi, B. Insights into a novel three-partner interaction between ants, coreids (Hemiptera: Coreidae) and extrafloral nectaries: implications for the study of protective mutualisms. *Arthropod-Plant Interact.* **11**, 525–536 (2017).
13. Clark, R. E., Farkas, T. E., Lichter-Marck, I., Johnson, E. R. & Singer, M. S. Multiple interaction types determine the impact of ant predation of caterpillars in a forest community. *Ecology* **97**, 3379–3388 (2016).
14. Feldhaar, H. et al. Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. *BMC Biol.* **5**, 48 (2007).
15. de Souza, D. J., Bézier, A., Depoix, D., Drezen, J. M. & Lenoir, A. *Blochmannia* endosymbionts improve colony growth and immune defence in the ant *Camponotus fellah*. *BMC Microbiol.* **9**, 29 (2009).
16. Gil, R. et al. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc. Natl Acad. Sci. USA* **100**, 9388–9393 (2003).
17. Kupper, M., Stigloher, C., Feldhaar, H. & Gross, R. Distribution of the obligate endosymbiont *Blochmannia floridanus* and expression analysis of putative immune genes in ovaries of the carpenter ant *Camponotus floridanus*. *Arthropod Struct. Dev.* **45**, 475–487 (2016).
18. Ramalho, M. O., Vieira, A. S., Pereira, M. C., Moreau, C. S. & Bueno, O. C. Transovarian transmission of *Blochmannia* and *Wolbachia* endosymbionts in the neotropical weaver ant *Camponotus textor* (Hymenoptera, Formicidae). *Curr. Microbiol.* **75**, 866–873 (2018).
19. Sauer, C., Dudaczek, D., Hölldobler, B. & Gross, R. Tissue localization of the endosymbiotic bacterium "*Candidatus* Blochmannia floridanus" in adults and larvae of the carpenter ant *Camponotus floridanus*. *Appl. Environ. Microbiol.* **68**, 4187–4193 (2002).
20. Sauer, C., Stackebrandt, E., Gadau, J., Hölldobler, B. & Gross, R. Systematic relationships and cospeciation of bacterial endosymbionts and their carpenter ant host species: proposal of the new taxon *Candidatus* Blochmannia gen. nov. *Int. J. Syst. Evol. Microbiol.* **50**, 1877–1886 (2000).
21. Wolschin, F., Hölldobler, B., Gross, R. & Zientz, E. Replication of the endosymbiotic bacterium *Blochmannia floridanus* is correlated with the developmental and reproductive stages of its ant host. *Appl. Environ. Microbiol.* **70**, 4096–4102 (2004).
22. Ward, P. S., Blaimer, B. B. & Fisher, B. L. A revised phylogenetic classification of the ant subfamily Formicinae (Hymenoptera: Formicidae), with resurrection of the genera *Colobopsis* and *Dinomyrmex*. *Zootaxa* **4072**, 343–357 (2016).
23. Sinotte, V. M., Freedman, S. N., Ugelvig, L. V. & Seid, M. A. *Camponotus floridanus* ants incur a trade-off between phenotypic development and pathogen susceptibility from their mutualistic endosymbiont *Blochmannia*. *Insects* **9**, 58 (2018).
24. Stoll, S., Feldhaar, H., Fraunholz, M. J. & Gross, R. Bacteriocyte dynamics during development of a holometabolous insect, the carpenter ant *Camponotus floridanus*. *BMC Microbiol.* **10**, 308 (2010).
25. Sameshima, S., Hasegawa, E., Kitade, O., Minaka, N. & Matsumoto, T. Phylogenetic comparison of endosymbionts with their host ants based on molecular evidence. *Zool. Sci.* **16**, 993–1000 (1999).
26. Degnan, P. H., Lazarus, A. B., Brock, C. D. & Wernegreen, J. J. Host–symbiont stability and fast evolutionary rates in an ant-bacterium association: cospeciation of *Camponotus* species and their endosymbionts, *Candidatus* Blochmannia. *Syst. Biol.* **53**, 95–110 (2004).
27. Extavour, C. G. & Akam, M. Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* **130**, 5869–5884 (2003).
28. Lehmann, R. & Nüsslein-Volhard, C. The maternal gene *nanos* has a central role in posterior pattern formation of the *Drosophila* embryo. *Development* **112**, 679–691 (1991).
29. Lehmann, R. Germ plasm biogenesis—an oskar-centric perspective. *Curr. Top. Dev. Biol.* **116**, 679–707 (2016).
30. Khila, A. & Abouheif, E. Reproductive constraint is a developmental mechanism that maintains social harmony in advanced ant societies. *Proc. Natl Acad. Sci. USA* **105**, 17884–17889 (2008).
31. Lynch, J. A. et al. The phylogenetic origin of *oskar* coincided with the origin of maternally provisioned germ plasm and pole cells at the base of the Holometabola. *PLoS Genet.* **7**, e1002029 (2011).
32. Lynch, J. A. & Roth, S. The evolution of dorsal–ventral patterning mechanisms in insects. *Genes Dev.* **25**, 107–118 (2011).
33. Akam, M. Hox genes, homeosis and the evolution of segment identity: no need for hopeless monsters. *Int. J. Dev. Biol.* **42**, 445–451 (1998).
34. Hughes, C. L. & Kaufman, T. C. *Hox* genes and the evolution of the arthropod body plan. *Evol. Dev.* **4**, 459–499 (2002).
35. Braendle, C. et al. Developmental origin and evolution of bacteriocytes in the aphid–*Buchnera* symbiosis. *PLoS Biol.* **1**, e21 (2003).
36. Matsuura, Y., Kikuchi, Y., Miura, T. & Fukatsu, T. *Ultrabithorax* is essential for bacteriocyte development. *Proc. Natl Acad. Sci. USA* **112**, 9376–9381 (2015).
37. Höhna, S. et al. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* **65**, 726–736 (2016).
38. Wernegreen, J. J., Degnan, P. H., Lazarus, A. B., Palacios, C. & Bordenstein, S. R. Genome evolution in an insect cell: distinct features of an ant–bacterial partnership. *Biol. Bull.* **204**, 221–231 (2003).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size. The experiments that were randomized are indicated below and investigators were not blinded to allocation during experiments and outcome assessment, with the exception of the qPCR experiment, in which the technician at the IRIC-Genomics Platform was blinded to allocation during experiments and outcome assessment.

Ant culturing and collection

Colonies were maintained in plastic boxes with glass test tubes filled with water constrained by cotton wool, and were fed a combination of mealworms, crickets, fruit flies and Bhatkar–Whitcomb diet[39]. All colonies were maintained at 25 °C, 70% relative humidity and a 12-h day:night cycle.

Colonies were collected from the following locations: *Aphaenogaster picea*, *Camponotus pennsylvanicus*, *Formica subscericea* and *Lasius niger* were collected at McGill Gault Nature Reserve (Quebec, Canada) (45° 32′ 12.4″ N, 73° 09′ 10.1″ W ± 1 km). *Camponotus novaeboracensis* ants were collected at Winnipeg (Manitoba, Canada) (49° 51′ 12.6″ N, 97° 08′ 14.0″ W ± 1 km). *Camponotus floridanus*, *Camponotus castaneus* and *Monomorium* sp. were collected at Gainesville (Florida, USA) (29° 42′ 05.7″ N, 82° 20′ 43.5″ W ± 1 km). *Colobopsis impressus* ants were collected at Gainesville (Florida, USA) (29° 41′ 07.4″ N, 82° 13′ 38.5″ W ± 1 km). *Camponotus ocreatus*, *Camponotus sansabeanus*, *Formica occulta* and *Veromessor pergandei* were collected at Miami (Arizona, USA) (33° 24′ 28.1″ N, 111° 00′ 14.5″ W ± 1 km). *Camponotus festinatus*, *Camponotus americanus*, *Camponotus sansabeanus*, *Brachymyrmex patagonicus*, *Nylanderia fulva* and *Nylanderia vividula* were collected at University of Texas at Austin, Brackenridge Field Laboratory (Texas, USA) (30° 17′ 2.40″ N, 97° 46′ 40.80″ W ± 1 km). *Myrmica americana* and *Prenolepis imparis* were collected at Medford (New York, USA) (40° 48′ 6.8566″ N, 73° 0′ 16.7756″ W ± 1 km). *Gigantiops destructor* ants were collected at ACTS Research Station (Maynas, Peru) (3° 14′ 60.00″ S, 72° 54′ 36.00″ W ± 1 km). *Anoplolepis gracillipes*, *Dolichoderus thoracicus*, *Oecophylla smaragdina*, *Paratrechina longicornis*, *Colobopsis leonardi* and *Polyrhachis rastellata* were collected at Mae Tang (Chiang Mai, Thailand) (location data not available). *Polyrhachis schlueteri* were collected at Bela Bela (Limpopo, South Africa) (24° 47′ 32.0″ S, 28° 17′ 30.6″ E ± 1 km). *Polyrhachis illaudata* and *Polyrhachis dives* were collected at Hong Kong region Guangdong (China) (location data not available). *Lasius emarginatus* ants were collected at Palmanova (Udine, Italy) (45° 54′ 31.5″ N, 13° 18′ 45.2″ E ± 1 km).

Aphaenogaster picea, *Camponotus pennsylvanicus*, *Formica subscericea*, *Lasius niger*, *Camponotus castaneus*, *Camponotus floridanus*, *Colobopsis impressus*, *Monomorium* sp., *Camponotus ocreatus*, *Camponotus sansabeanus*, *Formica occulta*, *Veromessor pergandei*, *Camponotus festinatus*, *Camponotus americanus*, *Camponotus sansabeanus*, *Brachymyrmex patagonicus*, *Nylanderia vividula*, *Myrmica americana* and *Prenolepis imparis* were collected by the laboratory of E.A. *Camponotus novaeboracensis* was collected by J. Rand, *Nylanderia fulva* was collected by E. Lebrun and *Gigantiops destructor* was collected by J. Gibson (laboratory of A. Suarez). *Anoplolepis gracillipes*, *Dolichoderus thoracicus*, *Oecophylla smaragdina*, *Paratrechina longicornis*, *Colobopsis leonardi* and *Polyrhachis rastellata* were purchased from Ants of Asia (P. Williams), and *Polyrhachis schlueteri*, *Polyrhachis illaudata*, *Polyrhachis dives* and *Lasius emarginatus* were purchased from Antstore (M. Sebesta).

Ovary dissections

This protocol was modified from a previous publication[40]. Ovaries were dissected in 0.1% PBSTween (1.86 mM NaH_2PO_4, 8.41 mM Na_2HPO_4, 1.75 M NaCl, 0.1% Tween20, pH 7.4) and kept on ice until fixation. First, the ovaries were removed from the oviduct. Ovaries were then separated into individual ovarioles, and the peritoneal sheath was then removed with fine forceps. Ovarioles were fixed in a solution of 5% formaldehyde (135 µl), 10% DMSO (100 µl) in 0.1% PBSTween (765 µl) for 25 min at room temperature, then washed with 0.1% PBSTween and gradually transferred to a solution of 100% methanol for storage.

Embryo collection and fixation

This protocol was modified from previous publications[40–42]. Embryos were treated with 4% hypochlorite solution (bleach) for 2 min. Embryos used for immunohistochemistry were then fixed using a 'slow formaldehyde fixing method' using PEMS (100 mM PIPES, 2 mM $MgSO_4$, 1 mM EGTA, pH 6.9) and were treated with proteinase K (New England Biolabs) in PBS at a final concentration of 0.08 U/ml. Embryos used for in situ hybridization were heat-fixed using a boiling hot solution of PBS-Triton (1.86 mM NaH_2PO_4, 8.41 mM Na_2HPO_4, 1.75 M NaCl, 0.03% Triton-X-100, pH 7.4).

Embryo staging

Timed egg depositions were collected in smaller setups and allowed to develop at 25 °C, 70% relative humidity along with a few workers, and fixed at two-hour windows. The embryos were DAPI-stained and observed under differential interference contrast (DIC) and wide-field fluorescence for staging. As far as possible, the staging scheme landmarks used correspond to Bownes' staging scheme for *Drosophila*[43].

Whole-genome shotgun sequencing

Whole DNA was isolated from 0–6-h-old embryos using Qiagen Genomic-tip 20/G kit. Shotgun sequencing was performed at Genome Quebec using Illumina HiSeq platform. Sequences were curated and BLAST searches performed using Geneious software[44].

Gene cloning and molecular biology

Gene sequences were obtained from NCBI GenBank database using genome BLAST against the assembled *C. floridanus* genome[45]. The accession numbers of genes used in this study are: *abdA* XM_020027891.2; *nos* XM_011266396; *osk* XM_011254572.2; *smg* XM_011254071.3; *stau* XM_011254361.3; *Ubx* XM_011259757.1; *en* XM_011252307.3. It was necessary to use a better-annotated *Ubx* cDNA sequence, which was submitted to GenBank under accession number MH801205. *Camponotus floridanus* and *L. niger* RNA was isolated using TRIzol (Invitrogen) from a pool of embryos and larvae of different developmental stages. RNA was then reverse-transcribed to synthesize a cDNA library. Specific primers were designed to amplify the gene fragments from cDNA libraries prepared from embryos and cloned in pGemT-easy vector (Promega) using standard procedures, and subsequently sequenced using Sanger sequencing at the Genome Quebec Innovation Centre. The primers used were: *osk* forward 5′-CGGAGAGCCTATTCCTTATC-3′, and reverse 5′-GCCAGAGATCTGATCCAATTA-3′, *nos* forward 5′-TCCCAGT TTGGACGAAGAATAAAG- 3′, and reverse 5′-GTTTTCCCGCAGAG TTTCTCAGTA- 3′, *stau* forward 5′-GCGAATTCACGGGTAGAGGT- 3′, and reverse 5′-GAAACACCAGCCGCATTCTG- 3′, *abdA* forward 5′-GTCTTC CTAAGAGCGACGAGC-3′, and reverse 5′-GTGGGTACCTTACTGACTGCC-3′, *Ubx* forward 5′-GCTTCTACGGAAGCCACCATC-3′, and reverse 5′-TGCTTC TCCTGCTCGTTTAGC-3′, *smg* forward 5′-TCACTTTTGCGTCGTCTACCT-3′, and reverse 5′-AGAGAGAGCCAGTTTGTGCC- 3′, *en* forward 5′-CGACACGAGCGAGGTATTGA- 3′, and reverse 5′-GAGGCCGATCGA TTTGACGA- 3′.

Identifying orthologues and paralogues

Amino acid sequence alignments were done using ClustalW in Geneious platform confirming the orthology of *vasa* (*vas*), *oskar* (*osk*), *nanos* (*nos*), *tudor* (*tud*), *germ cell-less* (*gcl*), *staufen* (*stau*), *caudal* (*cad*), *smaug* (*smg*), *wunen-2* (*wun2*), *aubergine* (*aub*), *heat shock protein 90* (*hsp90*), *argonaute 3* (*ago3*), *abdominal A* (*abdA*), *Ultrabithorax* (*Ubx*) and *engrailed* (*en*). The alignments are presented in Supplementary Figures 1–15. To search for any lineage-specific paralogues of the

germline genes that we studied, a blastn search was performed on the latest *C. floridanus* genome assembly accessed from: https://www.ncbi.nlm.nih.gov/assembly/1752781 using a maximum *E* value of 0.05; scoring of 2, −3; and gap cost of 5, 2. Only hits above an *e*-value cut off (e^{-20}) were considered and highlighted in bold. If the query subject had more than one hit but aligned to the same contig number, then it was concluded that no paralogues for the gene in question exist. However, if the query subject had more than one hit but was aligned to multiple contig numbers, then it was concluded that paralogues for the gene in question do exist. The hit tables are presented as Supplementary Table 3.

Immunohistochemistry and in situ hybridization

The following are primary antibodies we used in this study, the concentrations at which we used them at, and their source: mouse anti-HSP90 (1:100) antibody (BD bioscience 610418) and mouse anti-UbdA (1:4) antibody (FP6.87, DSHB), rabbit anti-Vasa (1:100) antibody (gift from P. Lasko), rabbit anti-Tudor (1:100) antibody (gift from P. Lasko), rabbit anti-Germ cell-less (1:300) antibody (gift from P. Lasko), rabbit anti-Aubergine (1:50) antibody (gift from P. Lasko), and rabbit anti-Oskar (1:100) antibody (gift from P. Lasko). Fluorescent secondary donkey anti-rabbit and anti-mouse polyclonal Alexa Fluor-488 (AbCam) antibodies were used at 1:500 dilution to detect the primary antibody, according a previous publication[30]. In situ hybridization was done according to previous publications[40,46], modified for in situ robot InsituPro VSi (Intavis) with the following modifications; the duration of wash steps was maintained according to the cited protocol but the buffer was exchanged every 5 min to increase agitation. Alkaline phosphatase secondary antibody anti-DIG-AP (Roche) was used to detect DIG-labelled probes and streptavidin-AP (Roche) reagent was used to detect biotin-labelled probes. Templates for probes were prepared using PCR with T7 and SP6 primers on the plasmids containing cloned gene fragments. Probe synthesis was done using SP6 or T7 RNA polymerase (Roche) according to the suppliers' directions. Probes were purified using phenol–chloroform and isopropanol precipitation method according to a previous publication[47], and used at 3 ng/µl final concentration. The probes consisted of 538 bp of *abdA* (bases 28–566), 424 bp of *nos* (bases 51–475), 848 bp of *osk* (bases 96–944), 987 bp of *stau* (bases 1121–2108) and 874 bp of *Ubx* (bases 32–906), 1,037 bp of *en* (bases 201–1237) and 949 bp of *smg* (bases 268–1197), in which base numbering starts at the start codon.

Microinjections and RNAi for phenotypic analysis

Embryos were collected as timed depositions from queens isolated with at least a dozen minor workers and at least six larvae and pupae. To eliminate any colony or day-of-injection related effects, embryos from multiple queens were collected, randomized between treatment and control and injected on the same day. Embryos were lined up alongside a fine glass capillary on a Petri dish lid lined with a thin layer of 2% agar in water and supplemented with 10 µl of 10 µg/ml ampicillin, modified from a previous publication[48]. Injection needles were prepared using a micropipette capillary puller. Microinjections were done using FemtoJet Express and InjectmanNI2 (Eppendorf) setup on a Zeiss Axiovert zoom inverted microscope using the following settings: control pressure 2 psi, injection pressure 18 ps, and injection time 0.1 s. The needle tip was broken open by gently pushing it against a glass coverslip immersed in halocarbon oil. The injection volume was adjusted after the needle was broken. Injection volumes were between 0.5 and 1 nl. Embryos were incubated at 25 °C 70% relative humidity chamber. Embryos were transferred every 24 to 48 h on to fresh 50-mm Petri-dishes containing 2% agar in water topped with a Whatmann filter paper and supplemented with 10 µl of 10 mg/ml ampicillin. DNA templates for double-stranded (ds) RNA were prepared using PCR with M13 forward universal primer and M13 reverse universal primer containing a T7 promoter overhang on plasmids containing cloned gene fragments as templates. The templates were used to generate dsRNA using T7 RNA polymerase (Roche) according to manufacturer's instructions. For controls, dsRNA was generated using the same method from a plasmid containing cloned 720 bp of the YFP coding sequence.

Quantitative PCR

Microinjections and RNAi. Embryos were collected as timed depositions from queens isolated with at least a dozen minor workers and at least six larvae and pupae. To eliminate any colony or day-of-injection related effects, embryos from multiple queens were collected, randomized between treatment and control and injected on the same day. Embryos were lined up alongside a fine glass capillary on a Petri dish lid lined with a thin layer of 2% agar in water and supplemented with 10 µl of 10 µg/ml ampicillin, modified from a previous publication[48]. Injection needles were prepared using a micropipette capillary puller. Microinjections were done using FemtoJet Express and InjectmanNI2 (Eppendorf) setup on a Zeiss Axiovert zoom inverted microscope using the following settings: control pressure 2 psi, injection pressure 18 psi and injection time 0.1 s. The needle tip was broken open by gently pushing it against a glass coverslip immersed in halocarbon oil. The injection volume was adjusted after the needle was broken. Injection volumes were between 0.5 and 1 nl. Embryos were incubated at 25 °C 70% relative humidity chamber. Embryos were transferred every 24 h on to fresh 50-mm Petri dishes containing 2% agar in water topped with a Whatmann filter paper and supplemented with 10 µl of 10 mg/ml ampicillin. DNA templates for dsRNA were prepared using PCR with M13 forward universal primer and M13 reverse universal primer containing a T7 promoter overhang on plasmids containing cloned gene fragments as templates. The templates were used to generate dsRNA using T7 RNA polymerase (Roche) according to manufacturer's instructions. For controls, dsRNA was generated using the same method from a plasmid containing cloned 720 bp of the YFP coding sequence.

Sample preparation for qPCR. Embryos were collected at stage 8, 5 days after injection and heat-fixed by immersing in a boiling hot solution of PBS-Triton (1.86 mM NaH_2PO_4, 8.41 mM Na_2HPO_4, 1.75 M NaCl, 0.03% Triton-X-100, pH 7.4) for 1 min followed by rinses with ice-cold PBS. Individual germline capsules, bacteriocytes and yolk sacs with intact germbands curled around them were separated using sharpened tungsten needles, and extraembryonic serosa tissue was discarded. For each gene, 40 individual samples were divided into 4 technical replicates of 10, and the 10 samples within each technical replicate were pooled and immediately placed in 200 µl Trizol reagent. RNA was prepared on the same day using standard Trizol method. First-strand cDNA synthesis was done using Superscript-II reverse transcriptase (ThermoFisher). Instead of universal oligo-dT primers an equimolar pool of the following 21 gene specific primers (including those of 8 endogenous control genes) was used to account for low yields in small tissue preparations: *vasa* 5′-CGATATCTGGTAGAAAGCCC-3′, *osk* 5′-GCCAGAGATCTGATCCAATTA-3′, *nos* 5′-GTTTTCCCGCAGAGTTTCTCAGTA- 3′, *tud* 5′-AGCGCCGGTTCTATCATGTC- 3′, *gcl* 5′-CCATCTCCAAGTATGTTCACC- 3′, *stau* 5′-GAAACACCAGCCGCATTCTG- 3′, *smg* 5′-AGAGAGAGCCAGTTTGTGCC- 3′, *ago3* 5′-TACACCCGTTATGCTTTTGA- 3′, *cad* 5′-AGAGGCGCCGATAGAGATGAA- 3′, *arm* 5′-TCTCGGTGCCTGTGATTCTG- 3′, *abdA* 5′-TCCAGGCCGCTTACGTGATG- 3′, *Ubx* 5′-TGCTTCTCCTGCTCGTTTAGC- 3′, *wun2* 5′-TCGTAATCGGTAGGTCGATGC- 3′, *act5c* 5′-GAACGGTGTTGGCGTACAGA- 3′, *tub* 5′-CGACGGAGAGTTGTTCGTGA- 3′, *argk* 5′-CCTGTCCAAGATCACCACCC- 3′, *ef1* 5′-AGTGGTCAATCCAGCAGGTG- 3′, *efl-like* 5′-GCAGCTGGTATTCCCGTTTG- 3′, *hisH-3* 5′-CCCTGAAAAGGGCCGATTGT- 3′, *rp60S* 5′-AACGTGCACTGGCATTTGTC- 3′, and *gadph* 5′-ATTCGCCATACGACGAGACC- 3′.

Quantitative PCR. Quantitative PCR was performed at IRIC-Genomics Platform using qPCR Taqman method[49] with the following primers: *vasa* forward 5′-CACAAGTACTTATTGTATCACCCACA-3′and

reverse 5′-GAAAATTTCTTGGCCTGTTGA-3′, *osk* forward 5′-AATCTCG TCGGAGAGCCTAT-3′and reverse 5′-AAATGCACGGAGACTCGAAA-3′, *nos* forward 5′-CCTTACCAACAGAATGCGTCT-3′and reverse 5′-TCCT TTAGCAGATGTTTTCGATAG-3′, *tud* forward 5′-ATTGTGGGTAC GAATATGTTATCG-3′and reverse 5′-ATGACAATGGTGTTAACATAAAGGAT -3′, *gcl* forward 5′-AAAACGATGGTTGGAAGTCAA-3′and reverse 5′-TGCCATTAAATCTGGTGCAA-3′, *stau* forward 5′-AACCCGCG AAACCATCTAT-3′and reverse 5′-CGTCACTTTTCTGGGTTTCG-3′, *ago-3*, forward 5′-TGGCATAGATGTCTATCATGCTG-3′and reverse 5′-GCAAC AAATCCTGCAACACTC-3′, *cad* forward 5′-ATGTCAATGCAGGCAGCAC-3′and reverse 5′-ACGTGGACGGAGATGTCG-3′, *wun2* forward 5′-TCTTGGC ACAATCGTAGCTTT-3′and reverse 5′-TCCGTGGAAGAATGCCTCT-3′, *tub* forward 5′-CACAGGCACGTATCGACAAC-3′and reverse 5′-GCCACGCG CATAATTGTT-3′, *act5c* forward 5′-CGTCATCAGGGTGTCATGG-3′ and reverse 5′-CAAGATACCTCTCTTCGATTGAGC-3′, *rp60S* forward 5′-GCGTTTCAAGGGCCAATAC-3′and reverse 5′-GCAGCATGTGA CGTGTTTTC-3′, *argk* forward 5′-TGGTAGACGCAGCGGTTT-3′and reverse 5′-AACGACTTGCTGTCGGATTC-3′, *efl-like* forward 5′-ACGTTATT GTCGAGGCCAAG-3′and reverse 5′-GGCAGGACGTATCTGCGTA-3′, *ef1* forward 5′-GCTGCAGTCGCATTTGTTC-3′and reverse 5′-ATCTTGGAAGAT GGCTCCAG-3′, *gapdh* forward 5′-GCGGTGCCAAGAAGGTTAT-3′and reverse 5′-CCAAGTTTACACCGACAACG-3′, *hisH-3*, forward 5′-CTACTAAAGCGG CGAGGAAG-3′and reverse 5′-CCAGGCCTATAACGATGAGG-3′.

The endogenous control genes used (the last 8 primer pairs above) were: *Actin5c*, *60S ribosomal protein*, *arginine kinase*, *efl-like*, *elongation factor1*, *gadph*, *histone H3* and *tubulin*. The most stable endogenous controls were established through the use of the algorithms integrated in the RefFinder package[50] that integrates four different protocols; GeNorm, BestKeeper, NormFinder and the comparative ΔC_T method[51–54] (Supplementary Fig. 16). Four of the endogenous controls (*gapdh*, *hisH3*, *rp60S* and *argk*) were deemed most stable and the geometric mean of these was used for calculating ΔC_T values for each target gene within each biological sample and replicate according to previously published recommendations[51] (Supplementary Fig. 16). Relative quantifications for *abdA* RNAi, *Ubx* RNAi and control *YFP* RNAi were calculated by the formula: relative quantification = $2^{-\Delta\Delta C_T}$, in which $\Delta\Delta C_T$ is the difference between ΔC_T in each RNAi sample and the average of ΔC_T values in all *YFP* RNAi replicates of that treatment group. $\Delta\Delta C_T$ values for each of the individual data points of the control *YFP* RNAi were also calculated using the average of all *YFP* RNAi from that particular biological replicate (black bars in Extended Data Fig. 4q, r). This method allows for consistency because the statistical analyses are performed on the same relative quantification values that are used to plot the bar graphs.

Antibiotic treatment

Two mature colonies were treated with rifampicin to test the effects of *Blochmannia* on embryonic development of *C. floridanus*. Rifampicin powder (Sigma; R883) was dissolved in water at a stock concentration of 2 mg/ml and then diluted 1:1 (final concentration 1 mg/ml rifampicin) in a 50% honey–water (Kirkland Signature) solution. Colonies were given fresh rifampicin–honey–water three times a week for two months. After two months, embryos were collected, fixed and stained with DAPI to confirm elimination of *Blochmannia*. Once elimination of *Blochmannia* was confirmed, embryos were collected and fixed for subsequent gene expression analysis. To rule out the possibility that the changes in phenotypes and gene expression or localization observed after rifampicin treatment are the unspecific effect of antibiotics were performed two controls: (1) a *C. floridanus* colony was treated with ampicillin, which does not eliminate *Blochmannia* from the colonies. Ampicillin powder (Fisher scientific; BP1760-25) was dissolved in water at a stock concentration of 400 mg/ml and then diluted 1:1 (final concentration 200 mg/ml ampicillin) in 50% honey–water solution. Colonies were treated in exactly the same manner as that for rifampicin. (2) An *L. niger* colony—a species that is in the same subfamily as *C. floridanus* but lacks *Blochmannia*—was treated with rifampicin in the same manner as *C. floridanus*. *Lasius niger* colonies were also treated with the same rifampicin regimen as *C. floridanus* and embryos were collected and fixed for subsequent gene expression analysis after at least two months.

Phylogenetic sampling, developmental characters and ancestral state reconstruction

Phylogenetic sampling. Thirty-one ant species were sampled in total: 26 from the subfamily Formicinae and 5 from 2 sister subfamilies of the Formicinae, the Myrmicinae (4 species) and the Dolichoderinae (1 species). Within the Formicinae, 14 in-group species within the Camponotini that evolved the obligate endosymbiosis with *Blochmannia* were sampled, and 12 out-group species were sampled that lack *Blochmannia*. Phylogenetic relationships and branch length information for these 31 species were obtained from previous molecular phylogenetic studies[26,55–57].

Developmental characters. The following five developmental characters were characterized for each species: (1) character 1 is defined as the presence of specific localization zones of mRNAs and proteins of the germline genes based on Vas protein. Character 1 has four states: an embryo with the presence of 1, 2, 3 or 4 localization zones of mRNAs and proteins of germline genes as illustrated in Fig. 4a; (2) character 2 is defined as the presence of specific localization zones of mRNAs and proteins of the maternal Hox genes *Ubx* and *abdA* based on the UbdA antibody that recognizes both Ubx and AbdA protein (with the exception of 1 species, *C. impressus*, which is based on *abdA* mRNA). Character 2 has four states: an embryo with the presence of 0, 1, 3 or 4 localization zones of mRNAs and proteins of maternal Hox genes *Ubx* and *abdA*, as illustrated in Fig. 4a; (3) character 3 is defined as the presence and type of obligate endosymbionts at the posterior of the egg on the basis of our own data and previous studies[8,9,58,59]. Previous phylogenetic evidence[10,25,38] showed that the three types of obligate endosymbionts within the Formicinae—the Camponotini obligate endosymbiont (*Blochmannia*), the Formica obligate endosymbiont and the Plagiolepidini obligate endosymbiont—were acquired independently and evolved convergently. Therefore, character 3 has four different states: an obligate endosymbiont at the posterior is absent; the Camponotini obligate endosymbiont (*Blochmannia*) is present at the posterior of the egg; the Formicini obligate endosymbiont is present at the posterior of the egg; or the Plagiolepidini obligate endosymbiont is present at the posterior of the egg, as illustrated in Fig. 4a; (4) character 4 is defined as the location of the embryo within the egg. Character 4 has two states: either the embryo is located in the posterior of the egg or the embryo is located in the anterior of the egg, as illustrated in Fig. 4a; (5) character 5 is defined as the germline capsule. Character 5 has two states: either the germline capsule is present or the germline capsule is absent, as illustrated in Fig. 4a.

Ancestral state reconstruction. RevBayes (v.1.7.10)[37] was used to reconstruct ancestral character states for the 5 developmental characters across 31 ant species sampled. RevBayes[37] uses Bayesian Markov chain Monte Carlo (MCMC) methods to estimate model parameters[60]. Ancestral states were estimated using two evolutionary models for discrete characters: the 'equal-transition rates' and 'unequal-transition rates' models. The equal-transition rates model assumes characters are equally likely to change from any one state to any other state, whereas the unequal-transition rates model assumes that the transition from any one state to any other state is unequal and can occur according to different rate parameters[37]. Both models were applied on each of the five developmental characters, and each model was run independently twice for 1,000,000 MCMC generations sampling every 500 generations. After completion of the MCMC analysis, the first 25% of the trees were discarded as a burn-in. Convergence between chains, likelihood scores and estimate sample size values were evaluated using Tracer

(version 1.7)[61]. The estimate sample size value for each parameter sampled from the MCMC analysis was always recorded as >1,000, indicating that the number of effectively independent draws from the posterior distribution from of all MCMC runs was adequate. Model selection was performed using marginal log-likelihoods, which represent the probability of the data given a specific model integrated over all possible parameter values[37]. Bayes factors were computed and used to estimate and compare the probabilities of the unequal and equal models given the data for each developmental character. Stepping-stone sampling (50 MCMC runs in RevBayes[37]) was used to approximate the marginal log-likelihoods[62]. The unequal model was found to be the model that best fit the data for all developmental characters (Extended Data Table 2). Nonetheless, the equal model also gives posterior probabilities similar to those of the unequal model (Extended Data Table 2), indicating that the reconstruction obtained for each dataset is robust to the evolutionary model assumed. Finally, we assessed the sensitivity of these posterior probabilities to the branch lengths obtained from the literature by repeating all of the above analyses, but setting all branch lengths equal to 1. The posterior probabilities obtained with all branch lengths equal to 1 were similar to those obtained from the literature (Supplementary Table 2), indicating that the reconstruction obtained for each dataset is robust to the branch lengths used.

Microscopy

We used a Zeiss Discovery V12 stereomicroscope and Zeiss Axiovision software to image embryos and ovaries. For high-resolution imaging, we used Leica SP8 confocal microscope. ImageJ2 was used for analysis of images[63].

Statistics and reproducibility

For a given gene, in situ hybridization and immunohistochemistry, the sample size for *C. floridanus* consisted of at least 30 embryos or ovarioles of similar stages; for other species that produce far fewer embryos, the sample size consisted of at least 5 embryos of similar stages. One hundred per cent of the embryos sampled showed the same expression patterns. In situ hybridization and immunohistochemistry experiments for *C. floridanus* were repeated at least eight times independently. For other species, these experiments were repeated at least four times. For RNAi experiments, phenotypes were considered reproducible if at least three independent replicates gave the same results. For qPCR, statistical analysis was performed using Graphpad Prism v7. or Microsoft Excel. Relative quantification values for *YFP* RNAi,s *abdA* RNAi and *Ubx* RNAi were calculated by the same method to ensure consistency between plotted results on the graph and for analysis of variance (ANOVA). Two-way ANOVA with replication was performed, in which *Ubx* RNAi was compared with *YFP* RNAi and *abdA* RNAi with *YFP* RNAi, treating RNAi as fixed and nine target genes as random effects. Each of the tissues (zone 1, zone2 and zone 3 + zone 4) was analysed by ANOVA as a separate experiment. The qPCR experiments were performed blind at the Genomic Platform facility at the Institute for Research in Immunology and Cancer. Fisher's exact test was performed to determine: (i) whether there is a significant difference in phenotype frequency (wild-type-like versus mild or severe) between control embryos collected from rifampicin-treated colonies versus tranplanted embryos collected from rifampicin-treated colonies. Analyses were considered statistically significant at $\alpha < 0.05$. For blinding and reproducibility, two different researchers independently performed the following steps without communicating each step: sample collection from colonies, randomization of embryos between treatments, treatment of samples, replicate maintenance, and data acquisition and analysis.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All relevant data are included in the Article, Extended Data and Supplementary Information. Raw sequence data that support the findings of this study have been deposited in GenBank with accession code MH801205, and in NCBI Sequence Read Archive with the accession code PRJNA625680. All raw image data that support the findings of this study are available in FigShare with the following identifiers: reference number 78072 (https://figshare.com/projects/The_origin_and_elaboration_of_a_major_evolutionary_transition_in_ants/78072); Fig. 1, https://doi.org/10.6084/m9.figshare.12133308; Fig. 2, https://doi.org/10.6084/m9.figshare.12133311; Fig. 3, https://doi.org/10.6084/m9.figshare.12133314; Fig. 4, https://doi.org/10.6084/m9.figshare.12133326; Extended Data Fig. 1, https://doi.org/10.6084/m9.figshare.12133296; Extended Data Fig. 2, https://doi.org/10.6084/m9.figshare.12133287; Extended Data Fig. 3, https://doi.org/10.6084/m9.figshare.12133110; Extended Data Fig. 4, https://doi.org/10.6084/m9.figshare.12133278; Extended Data Fig. 5, https://doi.org/10.6084/m9.figshare.12130902; Extended Data Fig. 6, https://doi.org/10.6084/m9.figshare.12131022; Extended Data Fig. 7, https://doi.org/10.6084/m9.figshare.12132993; Extended Data Fig. 8, https://doi.org/10.6084/m9.figshare.12131430. Source data are provided with this paper.

39. Bhatkar, A. & Whitcomb, W. Artificial diet for rearing various species of ants. *Fla. Entomol.* **53**, 229–232 (1970).
40. Khila, A. & Abouheif, E. In situ hybridization on ant ovaries and embryos. *Cold Spring Harb. Protoc.* **2009**, pdb.prot5250 (2009).
41. Rafiqi, A. M., Lemke, S. & Schmidt-Ott, U. *Megaselia abdita*: fixing and devitellinizing embryos. *Cold Spring Harb. Protoc.* **2011**, pdb.prot5602 (2011).
42. Rothwell, W. F. & Sullivan, W. in *Drosophila* Protocols (eds Sullivan, W. et al.) 141–157 (Cold Spring Harbor Laboratory Press, 2000).
43. Bownes, M. A photographic study of development in the living embryo of *Drosophila melanogaster. J. Embryol. Exp. Morpol.* **33**, 789–801 (1975).
44. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
45. Bonasio, R. et al. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**, 1068–1071 (2010).
46. Kosman, D. et al. Multiplex detection of RNA expression in *Drosophila* embryos. *Science* **305**, 846 (2004).
47. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1989).
48. Rafiqi, A. M., Lemke, S. & Schmidt-Ott, U. *Megaselia abdita*: preparing embryos for injection. *Cold Spring Harb. Protoc.* **2011**, pdb.prot5601 (2011).
49. Holland, P. M., Abramson, R. D., Watson, R. & Gelfand, D. H. Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl Acad. Sci. USA* **88**, 7276–7280 (1991).
50. Xie, F., Xiao, P., Chen, D., Xu, L. & Zhang, B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.* (2012).
51. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, research0034.0031 (2002).
52. Silver, N., Best, S., Jiang, J. & Thein, S. L. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol. Biol.* **7**, 33 (2006).
53. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **26**, 509–515 (2004).
54. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **64**, 5245–5250 (2004).
55. Borowiec, M. L. et al. Compositional heterogeneity and outgroup choice influence the internal phylogeny of the ants. *Mol. Phylogenet. Evol.* **134**, 111–121 (2019).
56. Blaimer, B. B. et al. Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. *BMC Evol. Biol.* **15**, 271 (2015).
57. Mezger, D. & Moreau, C. S. Out of South-East Asia: phylogeny and biogeography of the spiny ant genus *Polyrhachis* Smith (Hymenoptera: Formicidae). *Syst. Entomol.* **41**, 369–378 (2016).
58. Lilienstern, M. Beiträge zur Bakteriensymbiose der Ameisen. *Zeitschrift für Morphologie und Ökologie der Tiere* **26**, 110–134 (1932).
59. Jungen, H. Endosymbionten bei Ameisen. *Insectes Soc.* **15**, 227–232 (1968).
60. Pagel, M., Meade, A. & Barker, D. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**, 673–684 (2004).
61. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

62. Fan, Y., Wu, R., Chen, M.-H., Kuo, L. & Lewis, P. O. Choosing among partition models in Bayesian phylogenetics. *Mol. Biol. Evol.* **28**, 523–532 (2011).
63. Rueden, C. T. et al. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529 (2017).

Acknowledgements We thank L. Davis, R. Johnson, A. Suarez, J. Gibson, J. Rand, A. Wild and E. LeBrun for help with collecting ants; A. Vasquez-Correa, S. Joly, P. Ward and T. Oakley for help with ancestral-state reconstruction; M. Zayd and T. Chen for help with qPCR analysis; C. Metzl for translations; P. Lasko, S. F. Gilbert, J. Liebig, D. W. Wheeler, R. Rajakumar, C. Extavour, Y. Idaghdour, D. Schoen, A. Khila and members of the laboratory of E.A. for discussions or comments on the manuscript; and McGill University's Integrated Quantitative Biology Initiative (IQBI) and Advanced BioImaging Facility (ABIF) for imaging support. This work was supported by a doctoral fellowship from FQRNT (Quebec) to A.R., a Bezmialem Vakif University Grant (Turkey) to A.M.R., and an NSERC Discovery Grant and Steacie Fellowship (Canada), John Simon Guggenheim Fellowship (USA) and KLI Fellowship (Austria) to E.A.

Author contributions E.A., A.M.R. and A.R. conceived the project, designed experiments and collected ants. A.R. and A.M.R. performed all experiments. E.A. performed phylogenetic analyses. E.A., A.M.R. and A.R. wrote the manuscript.

Competing interests The authors declare no competing interests.

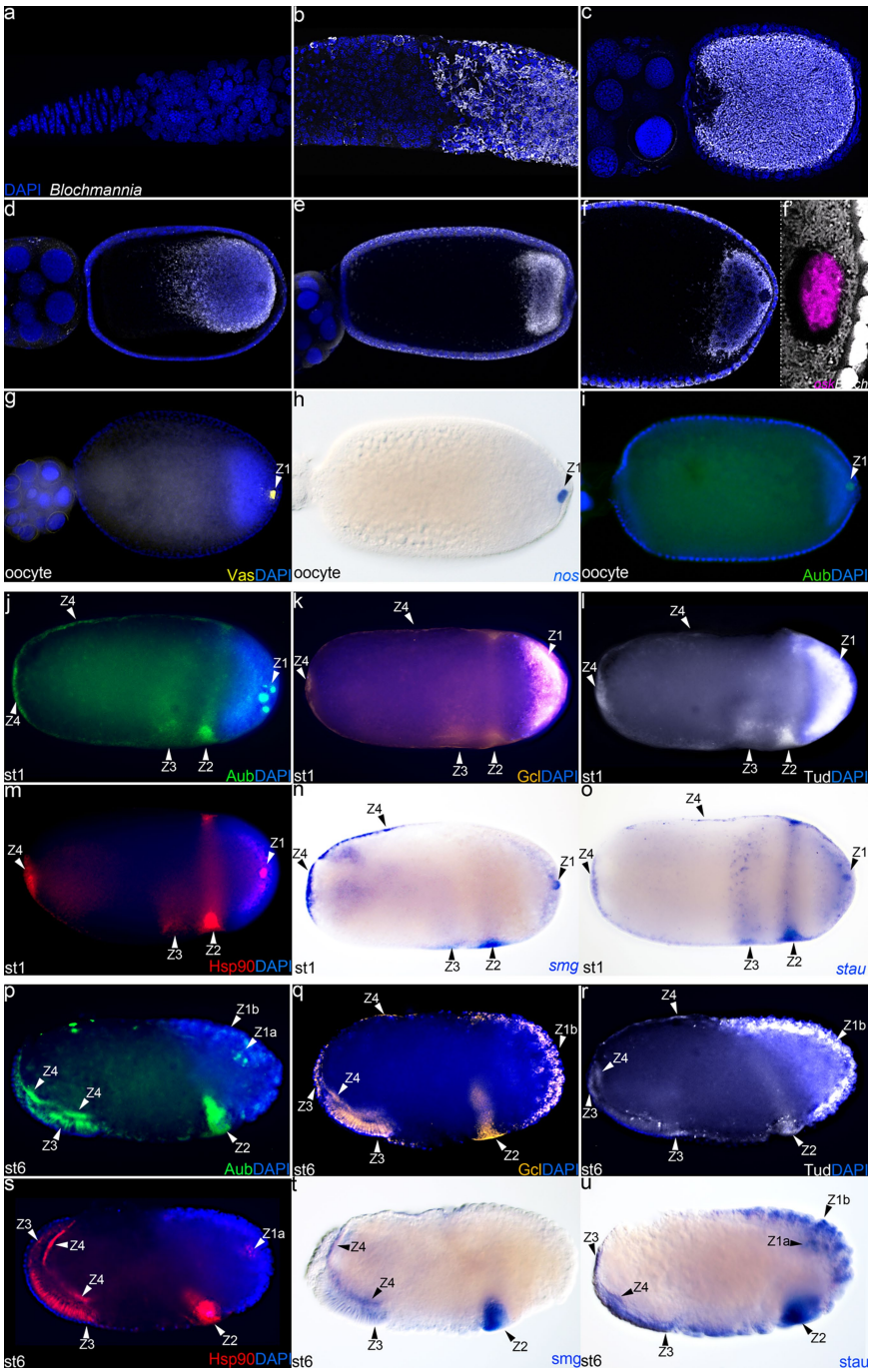
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2653-6.

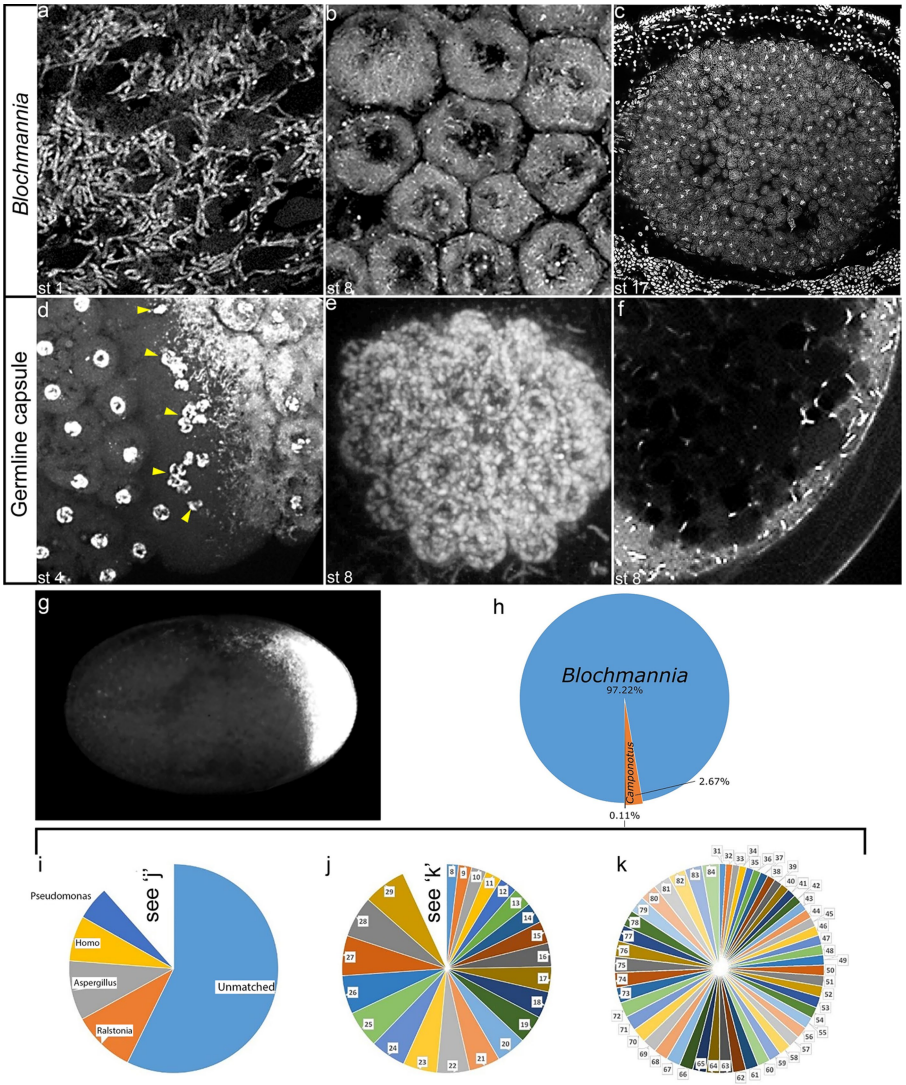
Correspondence and requests for materials should be addressed to E.A.

Peer review information *Nature* thanks Cameron R. Currie, Yannick Wurm and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

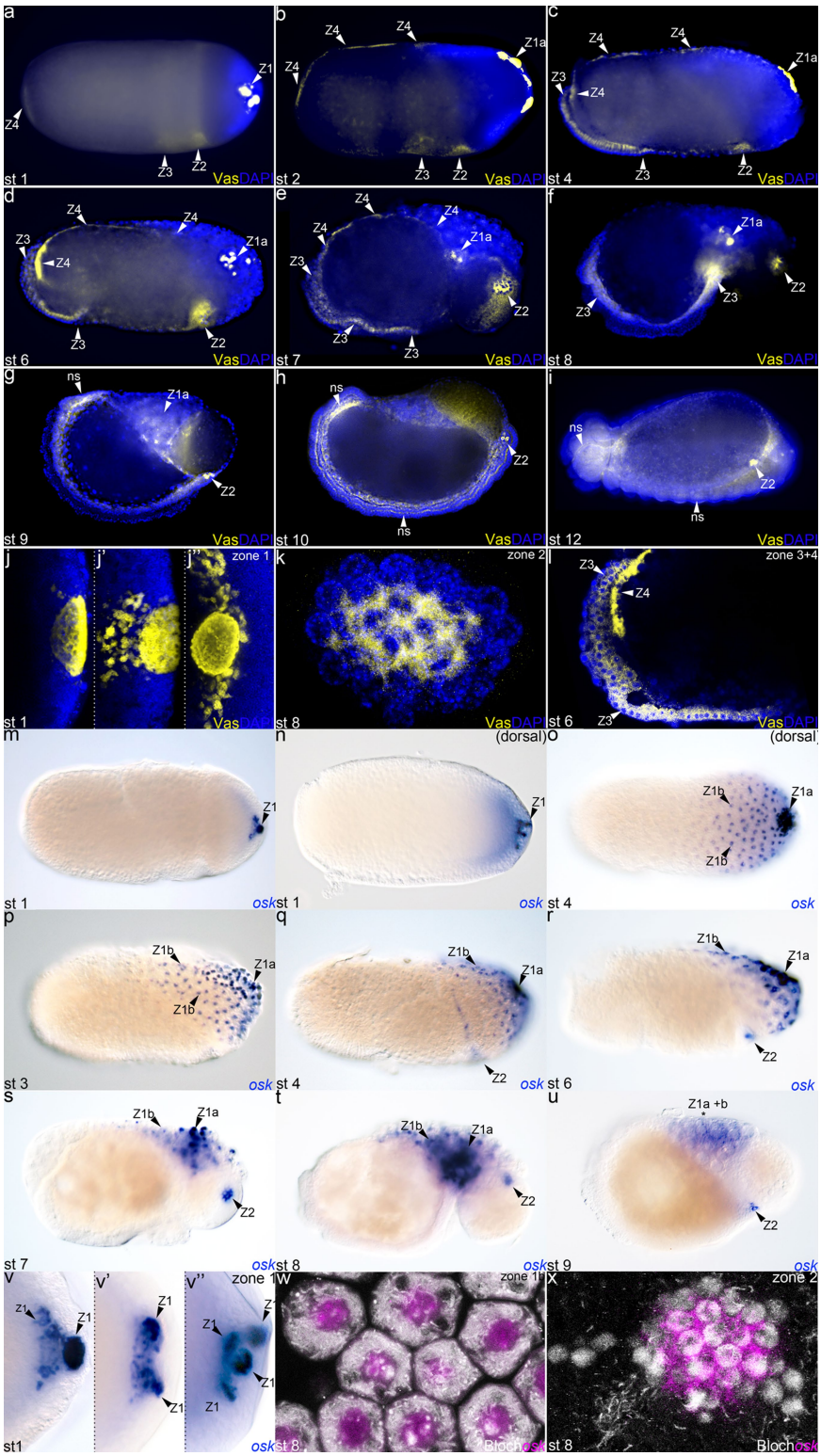
Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | Distribution of *Blochmannia* during oogenesis and the subcellular localization and expression of germline genes in *C. floridanus* oocytes and embryos. a–**f**, **f'**, Ovaries showing nuclear-stain DAPI in blue and *Blochmannia* in white: germline stem-cell niche without *Blochmannia* (**a**), germarium in which *Blochmannia* colonization occurs (**b**), *Blochmannia* initially fill the entirety of the cytoplasm of young oocytes (**c**) and progressively localize to the posterior pole of older oocytes (**d**–**f**), where *Blochmannia* surrounds the germplasm (**f'**). **f'**, **g**–**i**, Mature oocytes showing maternal expression of germline genes in oocytes, showing *osk* mRNA in magenta (**f'**), Vas protein in yellow (**g**), *nos mRNA* in blue (**h**), Aub protein in green (**i**) and nuclear-stain DAPI in blue. **j**–**o**, Subcellular localization zones in stage (st)-1 freshly laid eggs showing Aub protein in green (**j**), Gcl protein in orange (**k**), Tud protein in white (**l**), Hsp90 protein in red (**m**), *smg* mRNA in blue (**n**) and *stau* mRNA in blue (**o**). **p**–**u**, Expression in stage-6 cellular blastoderm embryos showing Aub protein in green (**p**), Gcl protein in orange (**q**), Tud protein in white (**r**), Hsp90 protein in red (**s**), *smg* mRNA in blue (**t**) and *stau* mRNA in blue (**u**). Arrowheads indicate subcellular localization or expression zones of germline genes: zone 1, zone 1a, zone 1b, zone 2, zone 3 and zone 4. Anterior is to the left, dorsal is to the top. In situ hybridization and immunohistochemistry experiments were repeated at least 8 times independently on $n \geq 30$ oocytes or embryos per developmental stage.

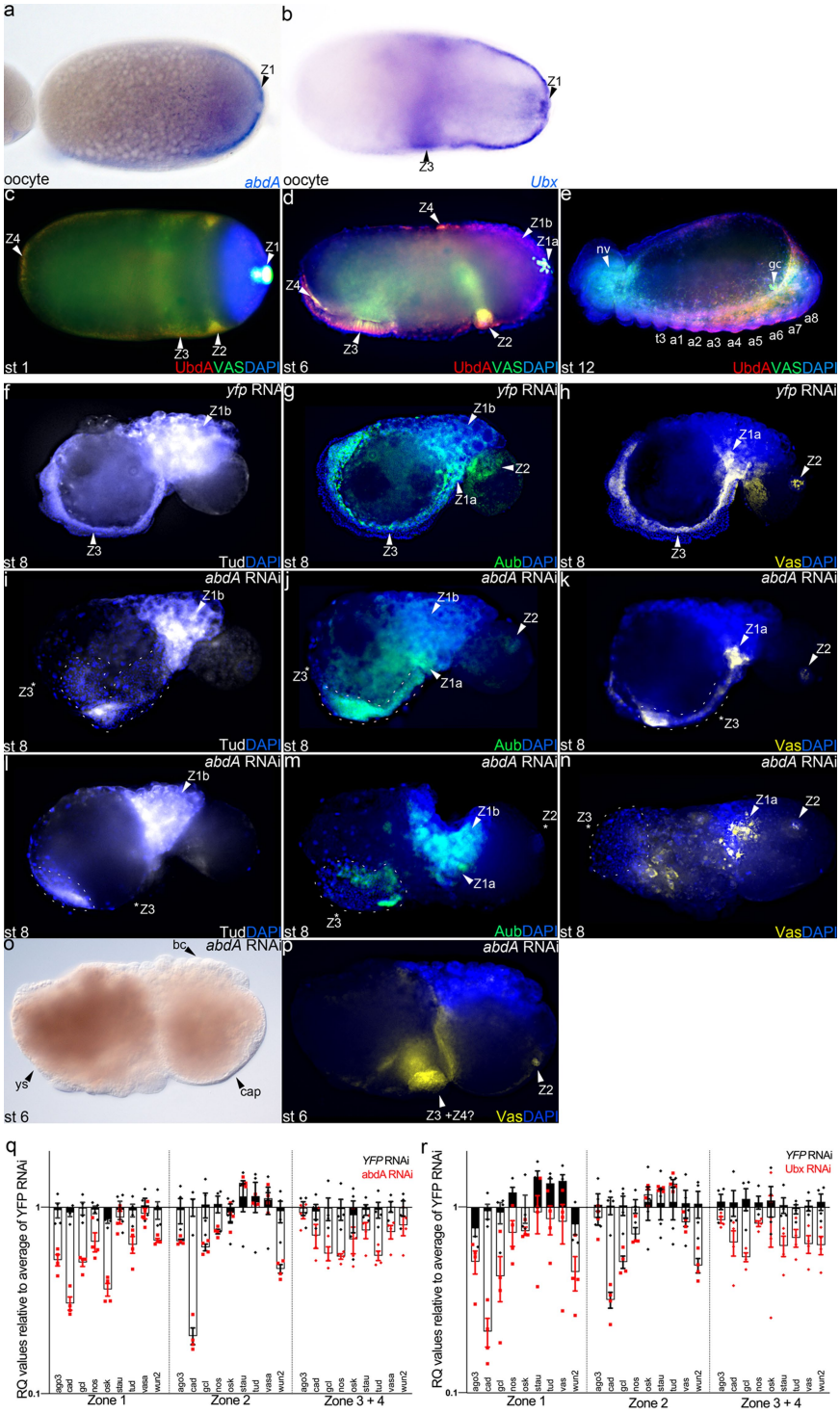


Extended Data Fig. 2 | *Blochmannia* segregates between bacteriocytes and germline capsule, and makes up 97.2% of DNA content in freshly laid eggs of *C. floridanus*. **a**–**f**, *Blochmannia*: at the posterior pole in freshly laid stage-1 eggs (**a**); inside bacteriocytes in stage-8 embryos (**b**); in bacteriocytes that line the midgut of stage-17 embryos (**c**); together with germline-precursor nuclei (yellow arrowheads) along the crest of the future germline capsule (**d**); surrounding the novel germline within the germline capsule (**e**); and as a small seed population for vertical transmission in the germline capsule (**f**). **g**, Freshly laid egg with DAPI in white, showing few zygotic nuclei in the anterior and *Blochmannia* at the posterior pole. **h**–**k**, Pie charts representing the number of Illumina Hi-Seq reads that match each of the indicated genera from DNA of freshly laid eggs. **h**, High abundance of *Blochmannia* DNA (blue) compared to that of host DNA (orange) and of other associated microorganisms (slim black slice) (shown in more detail in **i**–**k**) of decreasing abundance. We used a sequence similarity (*e*-value) of e^{-3} as a cut-off value for including any genus in our analysis. Numbers in **j** and **k** represent the following species: 8, *Serratia*; 9, *Leuconostoc*; 10, *Cupriavidus*; 11, *Cutibacterium*; 12, *Corynebacterium*; 13, *Mycobacterium*;14, *Candida*; 15, *Cyberlindnera*; 16, *Lactobacillus*; 17, *Brevibacterium*; 18, *Methylobacterium*; 19, *Pan*; 20, *Staphylococcus*; 21, *Sphingomonas*; 22, *Bradyrhizobium*; 23, *Plasmopara*; 24, *Bacillus*; 25, *Streptococcus*; 26, *Sphingopyxis*; 27, *Hyphomicrobium*, 28, *Acinetobacter*; 29, uncultured; 30, see **k**; 31, *Burkholderia*; 32, *Achromobacter*; 33, *Pichia*; 34, *Hyphopichia*; 35, *Penicillium*; 36, *Cyprinus*; 37, *Paenibacillus*; 38, *Brachybacterium*; 39, *Stenotrophomona*; 40, *Variovorax*; 41, *Streptomyces*; 42, *Sphingobium*; 43, *Nocardiopsis*; 44, *Dermabacter*; 45, *Sphingobacteriu*; 46, *Klebsiella*; 47, *Morganella*; 48, *Acidovorax*; 49, *Malassezia*; 50, *Lysobacter*; 51, *Rothia*; 52, *Pongo*; 53, *Rhodoplanes*; 54, *Microbacterium*; 55, *Rhodopseudomona*; 56, *Acheta*; 57, *Exiguobacterium*; 58, *Paraburkholderi*; 59, *Enterococcus*; 60, *Ramlibacter*; 61, *Actinomyces*; 62, *Bordetella*; 63, *Xanthomonas*; 64, *Brevundimonas*; 65, *Citrobacter*; 66, *Drosophila*; 67, *Lactococcus*; 68, *Mesorhizobium*; 69, *Candidatus*; 70, *Gluconobacter*; 71, *Rhodococcus*; 72, *Rubrivivax*; 73, *Saccharomyces*; 74, *Chelatococcus*; 75, *Hydrogenophaga*; 76, *Micrococcus*; 77, *Rhizobium*; 78, *Thauera*; 79, *Azospirillum*; 80, *Bosea*; 81, *Micromonospora*; 82, *Caulobacter*; 83, *Triticum*; 84, *Tsukamurella*. DAPI staining was repeated at least 4 times on $n \geq 30$ embryos per developmental stage.



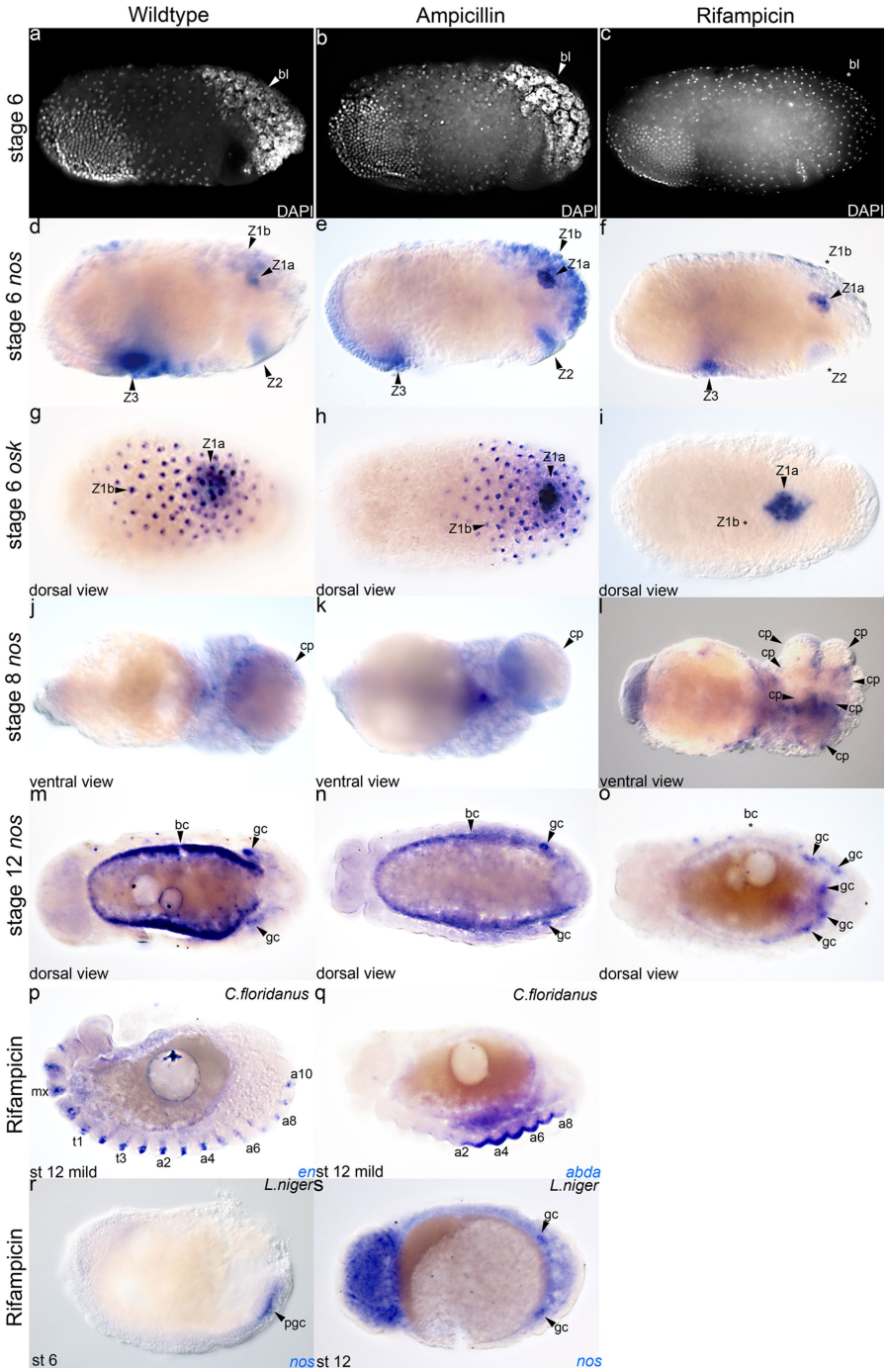
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Tracking the four functionally distinct zones through *C. floridanus* embryogenesis. a–**i**, Embryos showing Vas protein staining in yellow and DAPI in blue: freshly laid stage-1 and -2 eggs (**a**, **b**), cellular blastoderm stage-4 and -6 (**c**, **d**), gastrulation stage-7 (**e**), germband extension stage-8 to -10 (**f**–**h**) and segmentation stage-12 (**i**) embryos. **j**, **j′**, **j″**, **k**, **l**, Embryos showing higher-magnification confocal images of zone 1–4: freshly laid stage-1 egg, showing small germplasm foci budding off of the ancestral germplasm (**j**, **j′**, **j″**), stage-8 embryo showing novel germline (**k**), stage-6 embryo showing germband (zone 3) and yolk sac (zone 4) expression (**l**). ns, onset of Vas expression throughout the nervous system, brain and central nervous system in embryos from stage 9 onwards. **m**–**u**, *osk* mRNA in blue in stage-1 freshly laid egg (**m**, **n**), cellular blastoderm stage-3 and -4 embryo (**o**–**r**), gastrulation stage-7 embryo (**s**), and germband extension stage-8 and -9 embryo (**t**, **u**). **n**, **o**, Dorsal view, showing localization of small germplasm foci within the centre of bacteriocytes (zone 1b). **q**–**u**, Formation of the novel germline (zone 2). **u**, Embryo, showing loss of zone 1a and zone 1b. **v**, **v′**, **v″**, Small foci budding off the ancestral germplasm (zone 1). **w**, **x**, Higher-magnification confocal images of embryos, showing *osk* mRNA in magenta and DAPI in white. **w**, Stage-8 embryo, showing *osk* mRNA in magenta in the centre of bacteriocytes (zone 1b) surrounded by bacteria. **x**, Stage-8 embryo, showing expression of *osk* mRNA in the novel germline (zone 2). Zones are indicated with arrowheads. Anterior is to the left, dorsal is to the top. In situ hybridization and immunohistochemistry experiments were repeated at least 8 times independently on $n \geq 30$ embryos per developmental stage.

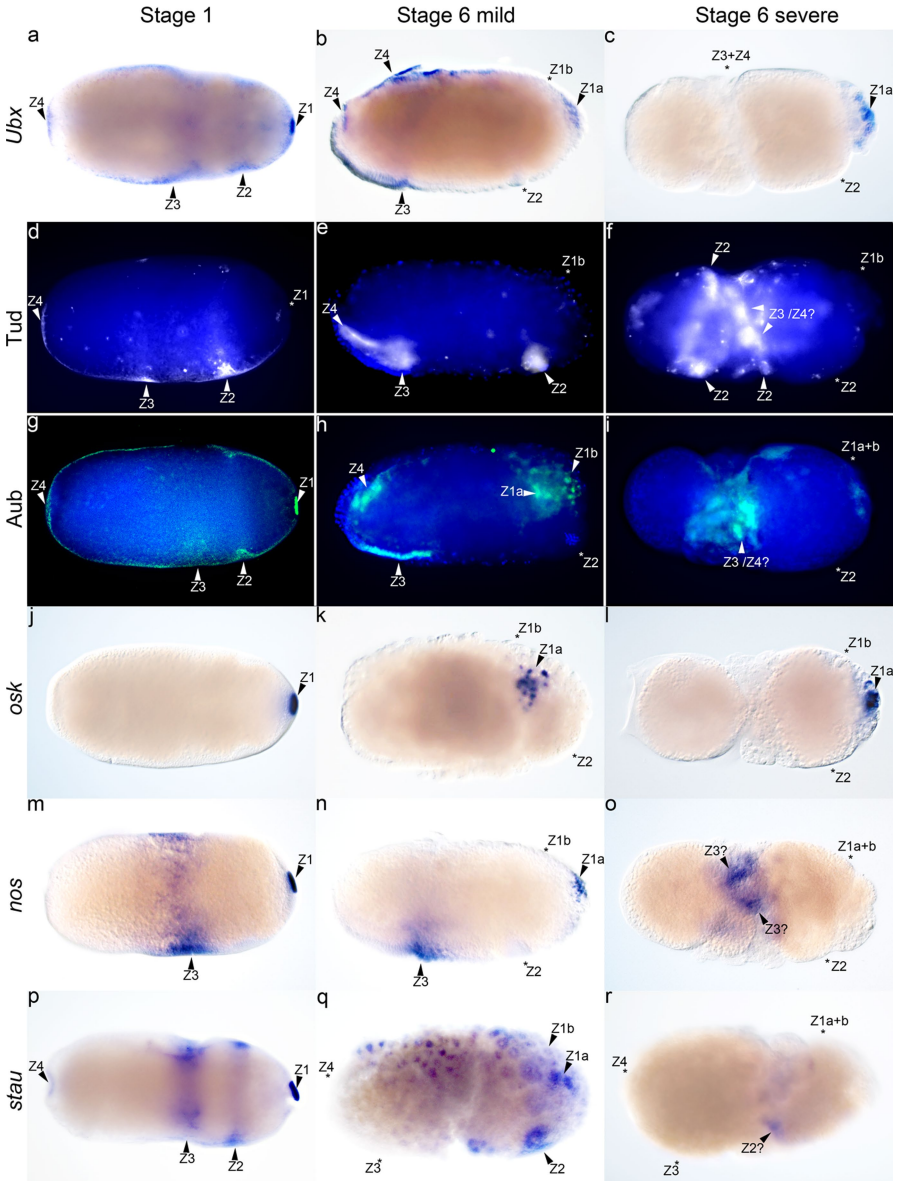


Extended Data Fig. 4 | See next page for caption.

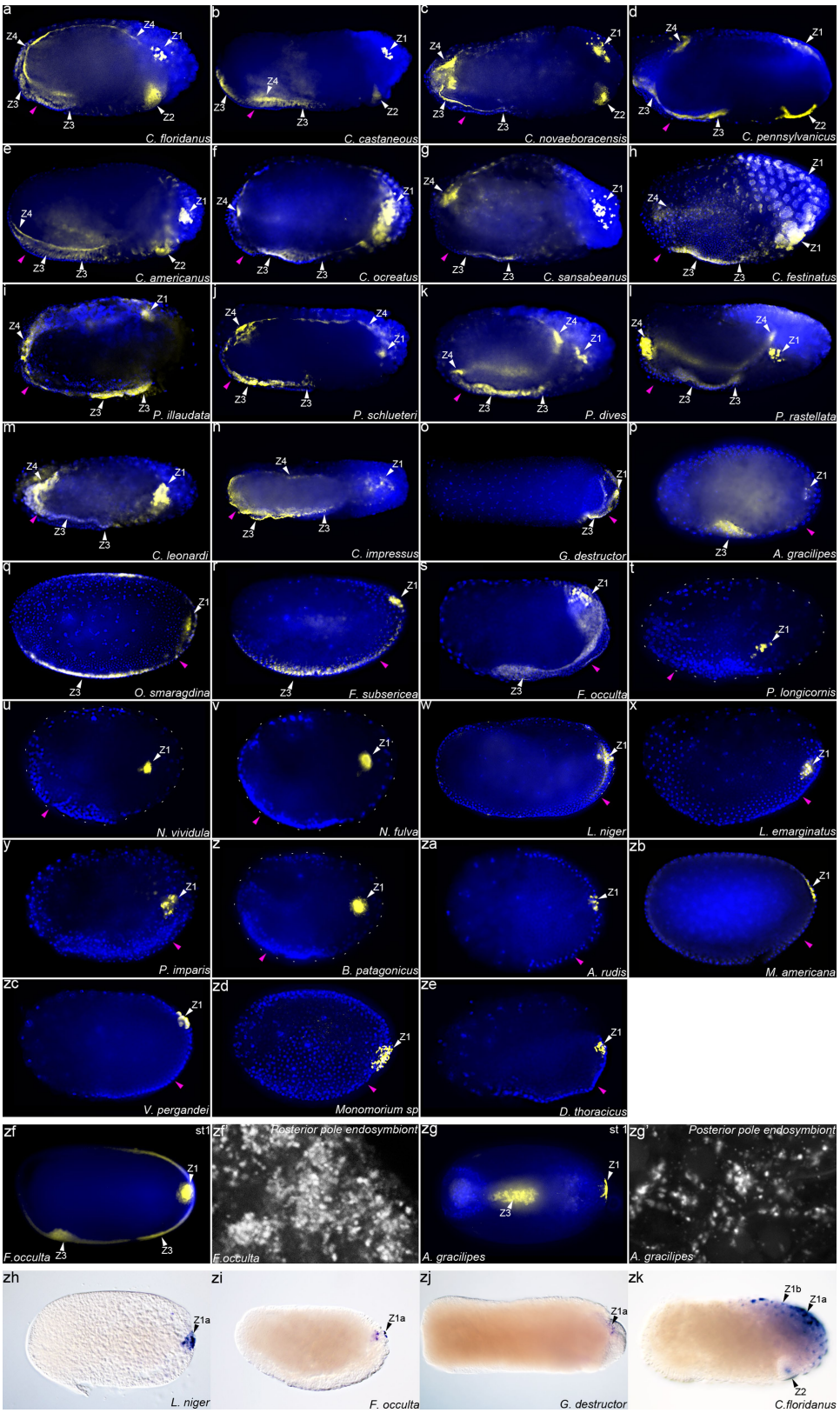
Extended Data Fig. 4 | *abdA* and *Ubx* are upstream of the germline genes. **a**, **b**, Mature oocytes stained for *abdA* mRNA (**a**) or *Ubx* mRNA (**b**) in blue. **c**–**e**, Colocalization (yellow and orange) of Ubx and AbdA (UbdA) protein in red, Vas protein in green and DAPI in blue in freshly laid stage-1 eggs (**c**), and stage-6 (**d**) and stage-12 (**e**) wild-type embryos. **f**–**p**, Expression of the germline genes in *YFP* RNAi ($n = 81$) (**f**–**h**) and high-concentration *abdA* RNAi embryos ($n = 61$ out 69) with DAPI in blue (**j**–**p**), stained for Tud in white (**f**, **i**, **l**), Aub in green (**g**, **j**, **m**) or Vas in yellow (**h**, **k**, **n**, **p**), and DIC of stage-6 embryo with severe phenotype (**o**). **i**–**k**, *abdA* RNAi embryos that are split along the midline ($n = 21$ out 61). **l**–**p**, Severe *abdA* RNAi phenotypes with an undifferentiated stub ($n = 34$ out of 61) (**l**–**n**) or in which the embryo is not detectable (**o**, **p**) ($n = 6$ out 61). Dotted outlines show changes in germband morphology and zone-3 expression after *abdA* RNAi. Zones are indicated with arrowheads. Asterisks indicate loss of germline gene expression within a specific zone. bc, bacteriocytes; cap, giant capsule; ys, yolk sac. Anterior is to the left, dorsal is to the top. **q**, **r**, Tissue-specific qPCR of nine germline genes (*x* axis; *ago3*, *cad*, *gcl*, *nos*, *osk*, *stau*, *tud*, *vasa* and *wun2*) from zone 1 (bacteriocytes), zone 2 (germline capsules), and zone 3 + zone 4 (embryonic germband + yolk sac) following *YFP* RNAi, low-concentration *abdA* RNAi and *Ubx* RNAi. Open bars represent mean relative quantification (RQ) values (*x* axis) and error bars represent standard error of the mean of: *abdA* RNAi (**q**) or *Ubx* RNAi (**r**). Black bars represent mean relative quantification values (*y* axis) and error bars represent standard error of the mean of *YFP* RNAi controls. Each individual data point (red squares) represents relative quantification value of a technical replicate from *abdA* or *Ubx* RNAi treatment relative to the average of all replicates of *YFP* RNAi control treatments (black diamonds) in that tissue. Two-tailed two-way ANOVA with replication for *abdA* RNAi versus YFP RNAi in zone 1 ($F = 129.311$, degrees of freedom (d.f.) = 1, $n = 54$, $P = 5.95504 \times 10^{-16}$ for *abdA* RNAi); zone 2 ($F = 20.733$, d.f. = 1, $n = 54$, $P = 3.04542 \times 10^{-5}$ for *abdA* RNAi); zone 3 + zone 4 ($F = 38.932$, d.f. = 1, $n = 54$, $P = 7.02605 \times 10^{-8}$ for *abdA* RNAi). Two-tailed two-way ANOVA with replication for *Ubx* RNAi versus YFP RNAi in zone 1 ($F = 66.278$, d.f. = 1, $n = 54$, $P = 5.84252 \times 10^{-11}$ for *Ubx* RNAi); zone 2 ($F = 12.628$, d.f. = 1, $n = 54$, $P = 0.000798519$ for *Ubx* RNAi); zone 3 + zone 4 ($F = 40.841$, d.f. = 1, $n = 54$, $P = 4.00577 \times 10^{-8}$ for *Ubx* RNAi). Raw data are in Source Data. In situ hybridization and immunohistochemistry experiments (**a**–**e**) were repeated at least 8 times independently on $n \geq 30$ embryos per developmental stage.



Extended Data Fig. 5 | Antibiotic treatment does not show unspecific effects. a–**c**, Stage-6 cellular blastoderm *C. floridanus* embryos with DAPI in white from wild-type colonies ($n \geq 30$ embryos) (**a**), colonies treated with ampicillin ($n \geq 15$ embryos) (**b**) or colonies treated with rifampicin ($n \geq 15$ embryos) (**c**). **d**–**o**, *C. floridanus* embryos stained for *nos* mRNA (**d**–**f**, **j**–**o**) and *osk* mRNA (**g**–**i**) in blue collected from wild-type colonies ($n \geq 30$ embryos each) (**d**, **g**, **j**, **m**), colonies treated with ampicillin ($n \geq 15$ embryos each) (**e**, **h**, **k**, **n**) or colonies treated with rifampicin ($n \geq 15$ embryos each) (**f**, **i**, **l**, **o**). **p, q**, Stage-12 mild-phenotype embryos collected from rifampicin-treated *C.-floridanus* colonies, showing expression of the segment polarity gene *en* in blue ($n \geq 15$ embryos) (**p**) or *abdA* mRNA in blue ($n \geq 15$ embryos) (**q**). **r**, **s**, *Lasius niger* embryos collected from rifampicin-treated colonies showing *nos* mRNA in blue in stage-6 embryos with normal primordial germ cells (pgc) ($n \geq 5$ embryos) (**r**) and stage-12 embryos with normal germ cells (gc) ($n \geq 5$ embryos) (**s**). Segments marked are as following: maxillary (mx), thoracic segments 1–3 (t1–t3) and abdominal segments 1–10 (a1–a10). White arrowheads indicate presence of *Blochmannia* (bl). White and black asterisks in embryos from rifampicin-treated colonies indicate loss of *Blochmannia* or loss of germline gene expression. **d**–**i**, Black arrowheads indicate zones. **j**–**l**, Black arrowheads indicate germline capsule(s) (cp). **m**–**o**, Black arrows indicate normal bacteriocyte (bc) and gonads (gc), development. Anterior is to the left, dorsal is to the top (**a**–**f**, **p**-**r**); dorsal is towards the reader in **g**–**i**, **m**–**o**, **s**; and ventral is towards the reader in **j**–**l**. In situ hybridization experiments were repeated at least 8 times (*C. floridanus*) or 4 times (*L. niger*) independently.

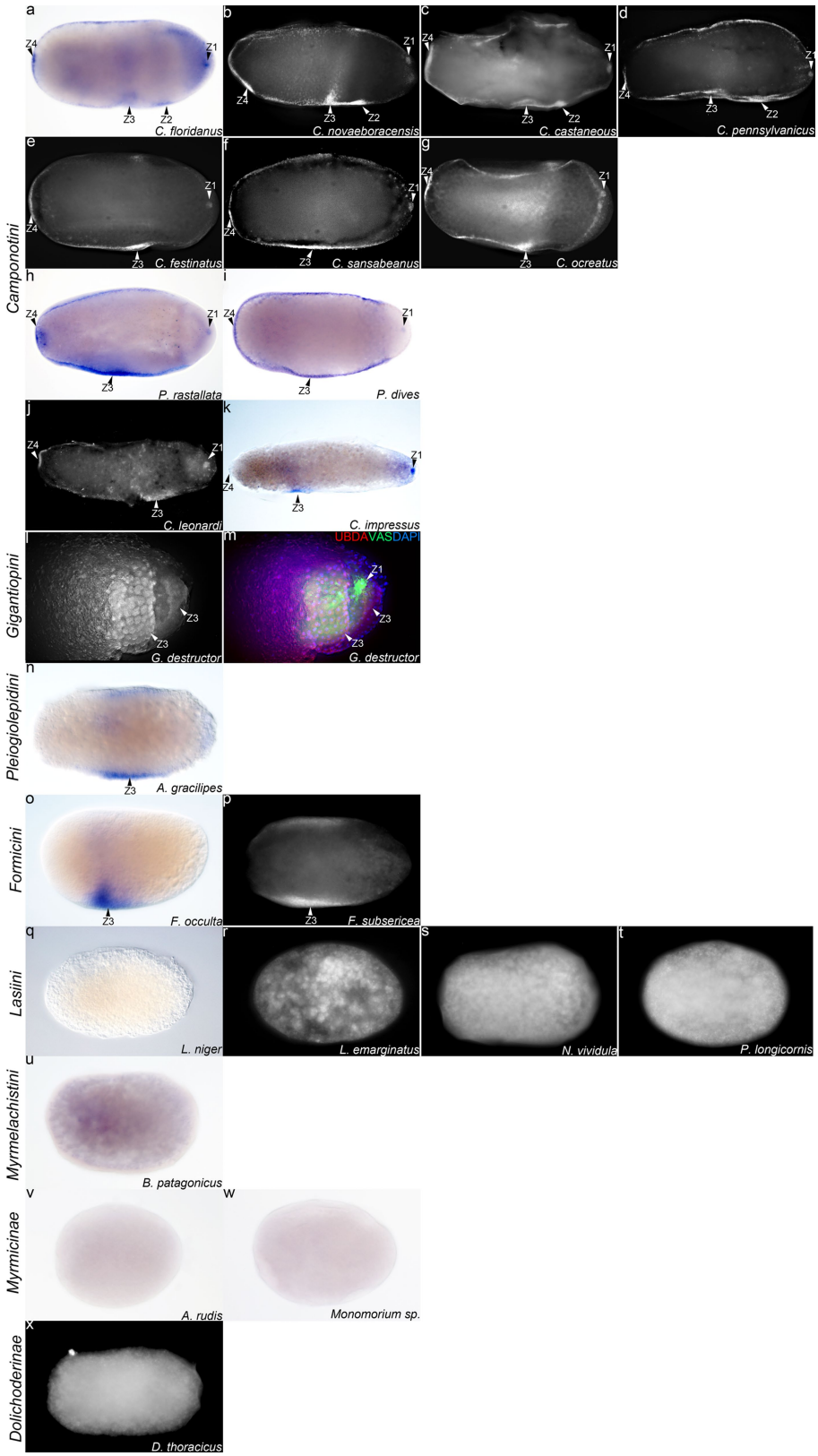


Extended Data Fig. 6 | *Blochmannia* maintains and selectively regulates mRNAs and proteins of maternal Hox and germline genes. a–**r**, Embryos from rifampicin-treated colonies stained for *Ubx* mRNA in blue (**a**–**c**), Tud protein in white (**d**–**f**), Aub protein in green (**g**–**i**), *osk* mRNA in blue (**j**–**l**), *nos* mRNA in blue (**m**–**o**) or *stau* mRNA in blue (**p**–**r**). **a**, **d**, **g**, **j**, **m**, **p**, Freshly laid stage-1 eggs showing no effect on the number of zones relative to wild type, except for in **d** Tud in white showing loss of zone 1 relative to wild type. **b**, **e**, **h**, **k**, **n**, **q**, Stage-6 mild-phenotype embryos with no observable morphological defects: asterisks indicate loss of specific mRNAs and proteins of Ubx and germline gene expression. **c**, **f**, **i**, **l**, **o**, **r**, Stage-6 severe-phenotype embryos showing morphological defects and loss or misexpression of germline and Hox genes. **d**–**i**, Fluorescent images with DAPI in blue. Zones of germline and Hox gene expression are indicated with arrowheads. Question marks indicate presumptive zones. Anterior is to the left, dorsal is to the top. In situ hybridization and immunohistochemistry experiments were repeated at least 4 times independently on $n \geq 15$ embryos per developmental stage.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Character states of germline localization zones, location of embryo, obligate endosymbiont and germline capsule. **a**–**z**, **za**, **zb**, **zc**, **zd**, **ze**, Cellular blastoderm stage embryos from Formicinae (**a**–**z**) and two sister subfamilies (**za**, **zb**, **zc**, **zd**) Myrmicinae and Dolichoderinae (**ze**), stained for Vas protein in yellow with DAPI in blue. **a**–**n**, Camponotini tribe. **a**, *Camponotus floridanus*. **b**, *Camponotus castaneous*. **c**, *Camponotus novaeboracensis*. **d**, *Camponotus pennsylvanicus*. **e**, *Camponotus americanus*. **f**, *Camponotus ocreatus*. **g**, *Camponotus sansabeanus*. **h**, *Camponotus festinatus*. **i**, *Polyrhachis illaudata*. **j**, *Polyrhachis schlueteri*. **k**, *Polyrhachis dives*. **l**, *Polyrhachis rastallata*. **m**, *Colobopsis leonardi*. **n**, *Colobopsis impressus*. **o**, Gigantiopini tribe: *Gigantiops destructor*. **p**, Pleigiolepidini tribe: *Anoplolepis gracilipes*. **q**, Oecophyllini tribe: *Oecophylla smaragdina*. **r**, **s**, Formicini tribe. **r**, *Formica subsericea*. **s**, *Formica occulta*. **t**–**y**, Lasiini tribe. **t**, *Paratrechina longicornis*. **u**, *Nylanderia vividula*. **v**, *Nylanderia fulva*. **w**, *Lasius niger*. **x**, *Lasius emarginatus*. **y**, *Prenolepis imparis*. **z**, Myrmelachistini tribe: *Brachymyrmex patagonicus*. **za**, **zb**, **zc**, **zd**, Myrmicinae. **za**, *Aphaenogaster rudis*. **zb**, *Myrmica americana*. **zc**, *Veromessor pergandei*. **zd**, *Monomorium* sp. **ze**, Dolichoderinae: *Dolichoderus thoracicus*. **zf**, **zg**, Freshly laid stage-1 eggs stained for Vas protein in yellow with DAPI in blue of *F. occulta* (**zf**) and *A. gracilipes* (**zg**). **zf′**, **zg′**, Endosymbiont at the posterior pole of *F. occulta* (**zf′**) and *A. gracilipes* (**zg′**). **zi**, **zj**, **zk**, **zl**, Cellular blastoderm stage embryos showing *osk* mRNA in blue, for *L. niger* (**zi**), *F. occulta* (**zj**), *G. destructor* (**zk**) and *C. floridanus* (**zl**). Zones of germline gene expression are indicated with white or black arrowheads. Magenta arrowheads indicate the location of the embryo within the egg. Experiments on all species were repeated 4 times independently on $n \geq 5$ embryos, except for *C. floridanus*, which was repeated 8 times independently with $n = 30$. Anterior is to the left, dorsal is to the top.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Character states of maternal Hox localization zones. **a**–**j**, **l**–**x**, Freshly laid stage-1 eggs from the Formicinae (**a**–**u**) and two sister subfamilies Myrmicinae (**v**, **w**) and Dolichoderinae (**x**) stained for UbdA (*Ubx* + *abdA* protein) in white or blue and (in **k**) *abdA* mRNA in blue. **a**–**k**, Camponotini tribe. **a**, *Camponotus floridanus*. **b**, *Camponotus novaeboracensis*. **c**, *Camponotus castaneous*. **d**, *Camponotus pennsylvanicus*. **e**, *Camponotus festinatus*. **f**, *Camponotus sansabeanus*. **g**, *Camponotus ocreatus*. **h**, *Polyrhachis rastallata*. **i**, *Polyrhachis dives*. **j**, *Colobopsis leonardi*. **k**, *Colobopsis impressus*. **l**, **m**, Gigantiopini tribe: *Gigantiops destructor*. In **m**, UbdA protein in red co-stained with Vas protein in green and DAPI in blue to distinguish germ cells from zone 3. **n**, *Pleigiolepidini* tribe: *Anoplolepis gracilipes*. **o**, **p**, *Formicini* tribe. **o**, *Formica occulta*. **p**, *Formica subsericea*. **q**–**t**, Lasiini tribe. **q**, *Lasius niger*. **r**, *Lasius emargiatus*. **s**, *Nylanderia vividula*. **t**, *Paratrechina longicornis*. **u**, Myrmelachistini tribe: *Brachymyrmex patagonicus*. **v**, **w**, Myrmicinae subfamily. **v**, *Aphaenogaster rudis*. **w**, *Monomorium* sp. **x**, Dolichoderinae subfamily: *Dolichoderus thoracicus*. Zones of maternal Hox localization are indicated with arrowheads: zone 1 (ancestral germline), zone 2 (novel germline), zone 3 (embryo) and zone 4 (anterior). Anterior is to the left, dorsal is to the top. Experiments on all species were repeated 4 times independently on $n \geq 5$ embryos, except for *C. floridanus*, for which experiments were repeated 8 times independently with $n = 30$ embryos.

Extended Data Table 1 | Combinatorial and dynamic localization or expression of germline and Hox genes across zones in stage-1 eggs and stage-6 embryos

		Wildtype stage 1				Wildtype stage 6				
#	Genes used	zone1	zone2	zone3	zone4	zone1a	zone1b	zone2	zone3	zone4
1	**Vasa**	+	+	+	+	+	-	+	+	+
2	***oskar***	+	-	-	-	+	+	+	-	-
3	***nanos***	+	-	+	-	+	+	+	+	-
4	**Tud**	+	+	+	+	-	+	+	+	+
5	**Aub**	+	+	+	+	+	+	+	+	+
6	***staufen***	+	+	+	+	+	+	+	+	+
7	***abdA***	+	-	+	-	+	+	+	+	+
8	***Ubx***	+	+	+	+	+	+	+	+	+
9	**Gcl**	+	+	+	+	-	+	+	+	+
10	**Hsp90**	+	+	+	+	+	-	+	+	+
11	***smg***	+	+	+	+	-	-	+	+	+

		Rifampicin stage 1				Rifampicin (mild) stage 6				
1	**Vasa**	+	+	+	+	+	-	-	+	+
2	***oskar***	+	-	-	-	+	-	-	-	-
3	***nanos***	+	-	+	-	+	-	-	+	-
4	**Tud**	-	+	+	+	-	-	+	+	+
5	**Aub**	+	+	+	+	+	+	-	+	+
6	***staufen***	+	+	+	+	+	+	+	-	-
7	***abdA***	-	-	+	-	-	-	-	+	+
8	***Ubx***	+	+	+	+	+	-	-	+	+

'+' indicates presence and '–' indicates absence. Grey shading indicates a change in localization or expression between embryos from wild-type and rifampicin-treated colonies.

Extended Data Table 2 | Posterior probabilities under the unequal- and equal-rates model for the five developmental characters

	Unequal model					Equal model				
Node #	germline localization zones	maternal Hox localization zones	obligate endo symbiont	Location of embryo	germline capsule	germline localization zones	maternal Hox localization zones	obligate endo symbiont	Location of embryo	germline capsule
0	0.95	0.91	0.95	0.96	1.00	1.00	0.96	1.00	0.92	0.98
1	0.96	0.91	0.96	0.96	1.00	1.00	0.96	1.00	0.92	0.98
2	0.97	0.92	0.97	0.97	1.00	1.00	0.95	1.00	0.95	0.98
3	0.98	0.95	0.98	0.98	1.00	1.00	0.95	1.00	0.97	0.98
4	0.99	0.95	0.99	0.99	1.00	1.00	0.93	1.00	0.99	0.98
5	0.95	0.90	0.95	0.95	1.00	1.00	0.95	1.00	0.92	0.98
6	0.89	0.85	0.94	0.95	1.00	0.95	0.89	1.00	0.92	0.98
7	0.98	0.96	0.98	0.93	1.00	1.00	0.95	1.00	0.88	0.98
8	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	0.99	0.98
9	1.00	0.98	1.00	0.74	1.00	1.00	0.96	1.00	0.68	0.98
10	1.00	0.98	1.00	0.61	1.00	1.00	0.96	1.00	0.62	0.98
11	1.00	0.98	1.00	0.97	1.00	1.00	0.96	1.00	0.97	0.98
12	0.89	0.71	0.85	0.96	1.00	0.97	0.73	0.99	0.92	0.98
13	0.99	0.97	0.95	1.00	1.00	1.00	0.96	0.98	0.99	0.98
14	0.90	0.72	0.85	0.96	1.00	0.98	0.74	0.99	0.92	0.98
15	0.90	0.73	0.82	0.95	1.00	0.98	0.75	0.98	0.92	0.98
16	0.88	0.72	0.80	0.93	1.00	0.96	0.74	0.97	0.89	0.98
17	0.92	0.92	0.96	0.96	1.00	0.98	0.94	0.99	0.95	0.98
18	0.98	0.98	0.99	0.99	1.00	1.00	0.97	1.00	0.99	0.98
19	0.97	0.96	1.00	1.00	1.00	1.00	0.97	1.00	1.00	0.98
20	1.00	0.99	1.00	1.00	1.00	1.00	0.97	1.00	1.00	0.98
21	1.00	0.94	1.00	1.00	1.00	1.00	0.95	1.00	1.00	0.98
22	1.00	0.99	1.00	1.00	1.00	1.00	0.98	1.00	1.00	0.98
23	0.96	0.95	1.00	1.00	0.98	1.00	0.97	1.00	1.00	0.97
24	0.92	0.91	1.00	1.00	0.95	0.98	0.94	1.00	1.00	0.95
25	0.96	0.95	1.00	1.00	0.98	1.00	0.97	1.00	1.00	0.97
26	0.91	0.90	1.00	1.00	0.94	0.98	0.93	1.00	1.00	0.94
27	0.98	0.98	1.00	1.00	1.00	1.00	0.98	1.00	1.00	0.98
28	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99
29	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99

The unequal-rates model is highlighted in grey shading. Figure 4a shows where the nodes are located on the phylogeny.

nature research

Corresponding author(s): Ehab Abouheif

Last updated by author(s): May 19, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Wide field fluorescence and DIC imaging of embryos was done using Zeiss Axiovision v4.9.1, Confocal miscroscopy was done using Leica SP8. qPCR data was collected by the Genomic Platform facility at the Institute for Research in Immunology and Cancer at the University of Montréal (Quebec, Canada) using a QuantStudio 7 Flex Real-Time PCR System.
Data analysis	DNA alignments, gene trees and BLAST analysis were performed on Geneious (R8). All statistical analyses were performed using Prism GraphPad v7 or Microsoft Excel (2013). Phylogenetic analyses were performed using RevBayes v1.7.1 and Tracer 1.7

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw sequence data that support the findings of this study have been deposited in Genbank with accession code MH801205 and in NCBI Sequence Read Archives with the accession codes [PRJNA625680, https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP256614].

All raw qPCR data has been provided as a Source Data File for Extended Data Figure 4q, r called 'Source Data Extended Data Fig. 4'

All raw image data that support the findings of this study are publicly available in figshare with the the following identifiers:
>>>Reference number 78072
https://figshare.com/projects/The_origin_and_elaboration_of_a_major_evolutionary_transition_in_ants/78072
Figure 1: 10.6084/m9.figshare.12133308
Figure 2: 10.6084/m9.figshare.12133311
Figure 3: 10.6084/m9.figshare.12133314
Figure 4: 10.6084/m9.figshare.12133326
Extended data figure 1: 10.6084/m9.figshare.12133296
Extended data figure 2: 10.6084/m9.figshare.12133287
Extended Data figure 3: 10.6084/m9.figshare.12133110
Extended data figure 4: 10.6084/m9.figshare.12133278
Extended data Figure 5: 10.6084/m9.figshare.12130902
Extended data figure 6: 10.6084/m9.figshare.12131022
Extended data figure 7: 10.6084/m9.figshare.12132993
Extended Data figure 8: 10.6084/m9.figshare.12131430

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	(1) For all statistical analyses, our ability to detect a significant difference between controls and experimental treatments at a significance level of alpha = 0.05 suggest that samples sizes used were sufficient. (2) For RNAi experiments for qPCR analysis: for both RNAi treatment and its control YFP RNAi, 40 individual dissected bacteriocytes were pooled into 4 replicates of 10 each. Similarly, 40 individual dissected capsules were pooled into 4 replicates of 10 each, and 40 individual dissected germbands and yolk sac were pooled into 4 replicates of 10 each. (3) For RNAi experiments for phenotype analysis: a minimum of 100 embryos was deemed sufficient for each gene targeted. The final sample size for each experiment was subject to variable mortality during the culturing process. The number of treated embryos (abdA or ubx RNAi) that exhibited phenotypes are reported as a number and percentage of the total number injected. We did the same for YFP RNAi control embryos. (4) For a given gene, in situ hybridization and immunohistochemistry sample size for C. floridanus consisted of at least 30 embryos or ovarioles of similar stages, while for other species that produce much fewer embryos, the sample size consisted of at least 5 embryos of similar stages. 100% of the embryos sampled showed the same expression patterns.
Data exclusions	(1) For all experiments, those embryos that died or were damaged during the experimental run or while handling post-experiment were excluded to ensure all developmental landmarks and tissues are consistently observable.
Replication	(1) For RNAi experiments for phenotype analysis: we performed each RNAi treatment (in parallel with its control YFP RNAi) at least three times independently. (2) For RNAi experiments for qPCR analysis: for both RNAi treatment and its control YFP RNAi, 40 individual dissected bacteriocytes were pooled into 4 technical replicates of 10 each. Similarly, 40 individual dissected capsules were pooled into 4 technical replicates of 10 each, and 40 individual dissected germbands and yolk sac were pooled into 4 technical replicates of 10 each. (3) In situ hybridization and immunohistochemistry experiments for C. floridanus were repeated at least eight times independently. For other species these were repeated at least four times. (4) For all experiments, all of our attempts at replication were successful.
Randomization	(1) To eliminate any colony or day-of-injection related effects in RNAi experiments for phenotype and qPCR analysis, embryos laid by multiple queens were collected, randomized between treatment and control, and injected on the same day. (2) For immunohistochemistry and in situ hybridization, embryos and ovarioles were collected from different colonies and were randomized before staining.
Blinding	(1) For qPCR, the experiments were performed blind at the Genomic Platform facility at the Institute for Research in Immunology and Cancer at the University of Montréal (Quebec, Canada). (2) Two different researchers independently performed the following steps without communicating each step: sample collection from colonies, randomization of embryos between treatments, treatment of samples, replicate maintenance, and data acquisition and analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Primary antibodies:
Rabbit anti-Vasa (1:100), Rabbit anti-Tudor (1:100), Rabbit anti-Germ cell-less (1:300), Rabbit anti-Aubergine (1:50), Rabbit anti-Oskar (1:100) (Source: gift from Paul Lasko Lab, Department of Biology, McGill University).
Mouse anti-HSP90 antibody (1:100) (Source: BD bioscience, Clone# 68, Cat # 610418)
Mouse anti-UbdA antibody (1:4) (Source: DSHB, Clone# FP6.87, Cat# UBX/ABD-A FP6.87, RRID:AB_10660834).

Fluorescent secondary antibodies:
Donkey anti-Rabbit polyclonal Alexa fluor-488 (1:500) (Source: AbCam, Cat# ab150073).
Donkey anti-Mouse polyclonal Alexa fluor-488 (1:500) (Source: AbCam, Cat# ab150105).

Alkaline phosphatase secondary antibodies:
anti-DIG-AP (1:500)(Source: Roche, Cat# 11093274910).
Streptavidin-AP (1:500)(Source: Roche, Cat# 11089161001).

Validation

All antibodies used in this study were validated using Drosophila samples and show conserved patterns of expression (see Supplementary Table 1). Furthermore, immunostains correspond directly with in situ hybridization stains of other germline genes, and for Ubx and abdA, the in situ hybridization stains matched the UbdA antibody (which recognizes both Ubx and abdA). Secondary antibodies did not require validation.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

The study did not involve laboratory animals

Wild animals

The study did not involve wild animals

Field-collected samples

(1) All species collected outside of Canada (see Methods) were imported into Canada under the following import permit numbers from the Canadian Food Inspection Agency: Arizona (P-2016-02919, P-2018-00739), New York (P-2016-02922, P-2018-00737), Texas (P-2016-02918, P-2018-00738), Florida (P-2016-02921, P-2018-00809), Thailand (Ants from Asia, P-2019-00011), Germany (Antstore, P-2019-00293).

(2) All colonies were housed in growth chambers at McGill University's Phytotron Facility under Plant Pest Containment Level 1 certification numbers PC-2016-057 and PC-2018-265 from the Canadian Food Inspection Agency.

(3) Colonies were maintained in plastic boxes with glass test tubes filled with water constrained by cotton wool, and were fed a combination of mealworms, crickets, fruit flies and Bhatkar-Whitcomb diet. All colonies were maintained at 25℃, 70% relative humidity and 12 h day:night cycle.

(4) None of the species collected are endangered as determined by their absence on the IUCN Red List of Threatened Species (https://www.iucnredlist.org/, and search Formicidae).

(5) None of the species were collected on protected lands. Therefore, none of the species collected in Canada, USA, Italy, and Thailand required collecting or export permits. For the species collected in South Africa and China, collecting or export permits were obtained by Antstore, and for species collected in Peru, collecting or export permits were obtained by Andrew Suarez (University of Illinois Ubana-Champaign).

(6) All experiments were performed on female ants.

(7) All colonies were collected from the wild as whole colonies or newly-mated queens using standard ant collecting procedures and

	subsequently transported in plastic containers. They are maintained alive indefinitely in growth chambers at McGill University's Phytotron Facility.
Ethics oversight	Ethics oversight was not required because ants do not require IRB or ethics approval

Note that full information on the approval of the study protocol must also be provided in the manuscript.