The genomes of non-bilaterian metazoans are key to understanding the molecular basis of early animal evolution. However, a full comprehension of how animal-specific traits, such as nervous systems, arose is hindered by the scarcity and fragmented nature of genomes from key taxa, such as Porifera. *Ephydatia muelleri* is a freshwater sponge found across the northern hemisphere. Here, we present its 326 Mb genome, assembled to high contiguity (N50: 9.88 Mb) with 23 chromosomes on 24 scaffolds. Our analyses reveal a metazoan-typical genome architecture, with highly shared synteny across Metazoa, and suggest that adaptation to the extreme temperatures and conditions found in freshwater often involves gene duplication. The pancontinental distribution and ready laboratory culture of *E. muelleri* make this a highly practical model system which, with RNAseq, DNA methylation and bacterial amplicon data spanning its development and range, allows exploration of genomic changes both within sponges and in early animal evolution.
One of the key events in the history of life was the evolutionary transition from unicellular organisms to multicellular individuals in which differentiated cell types work cooperatively\(^1\). In animals, the events that enabled this transformation can often be inferred by comparing the genomes of living representatives of non-bilaterian animals to those of bilaterians and their sister taxa, and determining shared characters and key differences between them\(^2\). However, the origins of several fundamental metazoan traits, such as tissues and nervous systems, are still unknown. Determining the origin of these characteristics requires more robust and contiguous genomic resources than are currently available for non-bilaterian animal taxa.

**Porifera**, commonly known as sponges, are one of the first lineages to have evolved during the rise of multicellular animals\(^3\) and are an essential reference group for comparative studies. The benchmark genome for sponges, *Amphimedon queenslandica*\(^4\), has provided a wealth of insight into the genomic biology of sponges\(^5,6\), yet studies of other sponge species have suggested that traits in *A. queenslandica* may not be representative of the phylum as a whole\(^7,8\). For example, its genome is one of the smallest measured in sponges\(^6\), it is highly methylated in comparison to other animals\(^9\), may have undergone some gene loss even in well-conserved families\(^10\) and has been described as possessing an ‘intermediate’ genomic state, between those of choanoflagellates and metazoans\(^5\). There are over 9200 species of sponge (http://www.marinespecies.org/porifera/\(^{11}\)), and understanding whether the unusual characteristics of *A. queenslandica* are typical of this large and diverse phylum can only be tested with additional, and more contiguous, genome assemblies.

Sponges diverged from the metazoan stem lineage in the Neoproterozoic\(^12\) and therefore are central to understanding the processes and mechanisms involved in the initial metazoan radiation. Sponges possess the fundamental characteristics shared by all animals, including development through embryogenesis to form tissues and signalling to coordinate whole body behaviour\(^13,14\). Most also have a highly conserved body plan, consisting of canals and pumping cells that filter water effectively\(^15\). However, within the four classes of sponges (Hexactinellida, Demospongiae, Calcarea, and Homoscleromorpha), several groups differ from this Bauplan. Glass sponges (Hexactinellida) have syncytial tissues and are the only sponges shown to propagate electrical signals\(^16\) while Cladorhizida (Demospongiae) are carnivorous and capture crustaceans with hook-like spicules\(^17\).

Despite the diversity of sponges in the most species-rich class, the Demospongiae, only one group, the Spongillida, made the transition to freshwater some 250–300 million years ago (Mya) (Fig. 1a), later diversifying into the extant range of modern taxa worldwide around 15–30 Mya\(^18\).

The transition to freshwater is one of the most remarkable evolutionary trajectories marine animals can undergo, as it requires a complete spectrum of physiological adaptations to novel habitats. Not only can freshwater sponges, which consist of a single layer of cells over a scant extracellular matrix, control their osmolarity in freshwater\(^19\), they can withstand extremely cold temperatures and even freezing, as they inhabit some lakes that see temperatures below \(-40^\circ\)C\(^20\) and can also tolerate extreme heat and desiccation in desert sand dunes and high up on tree trunks\(^21,22\). Freshwater sponges are both unfamiliar and yet so common worldwide that under the right conditions they can foul drinking water reservoirs, waste treatment plants, intake pipes, and cooling systems for power plants\(^23\). The main adaptation which permits colonisation of such extreme habitats is the production of sophisticated structures called gemmules, a distinctive stage in the life of these sponges\(^21\). The events that allowed colonisation of freshwater which are required for adaptation to extreme habitats by sponges are not yet fully understood\(^24\). Whether genomes of freshwater species are remarkably changed from those seen in marine sponges and other animals is also yet to be investigated.

The freshwater sponge *Ephydatia muelleri* (Lieberkühn, 1856) (Fig. 1) is found in rivers and lakes throughout the northern hemisphere (Supplementary Note 1, Supplementary Table 1). Because of its global distribution and century-long history of study both in situ and in the laboratory, *E. muelleri* is an outstanding model for asking questions about adaptation and the evolution of animal characters. It has separate sexes, allowing the study of inheritance\(^25\), but more practically, gemmules are clones that can easily be cultured at room temperature\(^26–28\). They also tolerate freezing\(^29\); this species can be stored at \(-80^\circ\)C for several years prior to hatching\(^30\).

Here, we present a chromosome-level assembly of the 326 million base pair (megabases, Mb) *Ephydatia muelleri* genome. The highly contiguous assembly of this sponge genome is an exceptionally rich resource that reveals metazoan typical regulatory elements, macrosynteny shared with other non-bilaterians and chordates, and allows analysis of structural chromatin variation across animals. The high gene count of sponges compared to other animals is shown to be a feature of gene duplication. These analyses, together with the evaluation of RNA expression in development and host–microbe relationships across the range of this species provide key data for understanding the genomic biology of sponges and the early evolution of animals.

## Results and discussion

### A chromosomal-level genome, with higher gene content than most animals

We have produced a high-quality assembly of the 326 Mb *Ephydatia muelleri* genome using PacBio, Chicago, and Dovetail Hi-C libraries sequenced to approximately 1490 times total coverage (Supplementary Note 2). The resulting assembly has 1444 scaffolds with a scaffold N50 of 9.8 Mb (Fig. 2a–c, Supplementary Note 3) and 83.7% of the genome (270 Mb) is encompassed in the largest 24 scaffolds in the assembly. These 24 scaffolds encompass 22 of the 23 *Ephydatia muelleri* chromosomes (2n = 46)\(^32\) as single sequences, with one chromosome represented by two sequences likely split at the centromere (Fig. 2b).

One scaffold (the 25th largest), containing the partial genome of a member of a *Flavobacterium* spp. (3.09 Mb with 3811 genes), a possible symbiont species, was removed from the final assembly and analyzed separately (Supplementary Note 4, Supplementary Table 1). The remaining 1419 scaffolds (53.3 Mb) contain less than 16% of the genome and have no other clear bacterial sequence content (section 3.5 of Supplementary Note 3).

The *E. muelleri* assembly size corresponds well to the predicted genome size based on Feulgen image analysis densitometry (0.34 pg) and flow cytometry (0.33 pg)\(^30\), and is twice the size of the well-studied *A. queenslandica* genome (166 Mb; Supplementary Table 2)\(^4\), but similar in size to *Sycon ciliatum* (genome size 357 Mb)\(^31\). The genome is approximately 43% G+C (Supplementary Table 2: cf. *A. queenslandica*: 35.82%, *S. ciliatum*: 46.99%). Nearly 47% of the genome is repetitive sequences, compared to *A. queenslandica* 43% and *S. ciliatum* 28% (Supplementary Table 6). The *E. muelleri* genome browser and other resources are available at https://spaces.fasci.ualberta.ca/ephybase/.

Sponges, perhaps counterintuitively, have more genes than most other animals, and almost twice the number of genes found in humans. The *E. muelleri* genome contains 39,245 predicted protein-coding loci (Supplementary Table 10). This number of genes compares well with recent estimates for *A. queenslandica* (40,122)\(^5\), *S. ciliatum* (32,309)\(^31\), and *Tethya wilhelma* (37,416)\(^8\). The gene annotations contain 90.10% of the 303 eukaryotic
**Fig. 1 Ephydatia muelleri biology and inter-relationships.** a Diagrammatic cladogram of poriferan inter-relationships, showing fossil-calibrated divergence clock at bottom and at key nodes. These estimates of divergence were taken from prior analyses as shown in Supplementary Fig. 1. Tree is rooted with choanoflagellate species. Key clades and previously sequenced species are boxed and indicated. b Light microscope images of E. muelleri in culture and in situ (photos: S. Leys, scale 1 mm in left image, 1 cm at right). c Life cycle of E. muelleri, showing asexual (left) and sexual (right) modes of reproduction. Gemmules indicated as yellow dots in adult tissue. Asexual reproduction proceeds via hardy gemmules produced internally (i), which are separated from adults (ii), hatch and directly develop (iii) into adult tissue (iv). Sexual reproduction proceeds via generation of gametes (i*), embryonic development into a mobile parenchymella larva (ii*), and development as an independent adult (iii*).

**Fig. 2 Genome assembly and architecture.** a Hi-C contact map, visualised in Juicebox.js, with number of contacts coloured in blue according to the scale below. b Histogram of assembly scaffold sizes and corresponding chromosomes drawn to scale after Ishijima et al. Note the large difference in size between chromosome 1 and the additional 22 chromosomes, which was reflected in our assembly. c Representation of the relative completeness of the E. muelleri assembly compared to those of a number of commonly used genomes. N50 is represented by diameter of circles and number of scaffolds / genome size are on the x and y axes, respectively. The high level of contiguity of our assembly is obvious in this comparison. d Conserved syntenic signal in Ephydatia muelleri with Trichoplax adhaerens, e and with Branchiostoma floridae. Each small green dot on the matrices represents an identified homologous gene pair. Individual scaffolds for each species pair possess numerous homologous genes. Red-boxed locations are shown in detail to the right of each matrix as chromosomal representations. Each black line on the chromosome images represents a gene linked in green to the homologous gene in the other species. It is clear that numerous genes are conserved between ancestrally homologous chromosomal regions between these species, although their relative arrangement within chromosomes is shuffled, as expected under the DCJ-DS model of syntenic evolution. Source data are provided as a Source Data file.
BUSCO genes (83.83%, [254 genes] as complete models, section 3.2 of Supplementary Note 3). Approximately 74% of the E. muelleri proteins (29,571) have similarity to other organisms as determined by DIAMOND BLASTx (nr database, E-value threshold of $10^{-5}$), and nearly half of these hits (13,579) were best-matches to other sponge genes (section 5.2 of Supplementary Note 5, Supplementary Data 2). We assigned 19,362 genes (approximately 50% of the total number) a full functional annotation using BLAST2GO (section 5.3 of Supplementary Note 5). Introns size and intergenic distance scale with genome size and accordingly both intron size and intergenic distance are greater in E. muelleri compared to A. queenslandica (Supplementary Table 10, Supplementary Data 3), but these values are still relatively small when compared to other animal genomes.

The abundance of genes seen in E. muelleri is in part due to tandem duplication. Many gene clusters have identical intron–exon structure between duplicated genes, suggesting that the mechanism of gene duplication is from replication slippage and unequal crossing-over. E. muelleri also shows evidence of widespread segmental duplication, with many gene clusters replicated. For example, both scaffold_0002 and scaffold_0004 contain a large cluster of predicted homologues of integrins, while on scaffold_0004 the integrin cluster overlaps with a large cluster of 177 predicted E3-ubiquitin ligases (e.g., Supplementary Fig. 6). These cluster duplications are recognisable by their close similarity in sequence, especially in coding sequences, but intergenic and intronic sequences are highly variable, strongly suggesting that these clusters are true duplications and not assembly artifacts. Even BUSCO genes, which are found in single copy in most genomes, are duplicated in E. muelleri with 19.6% represented by more than one copy (section 3.2 of Supplementary Note 3, Supplementary Tables 4, 5).

Conservation of synteny with other metazoans. Sponges diverged from other metazoans in the Neoproterozoic (540–1000 mya) and yet we found evidence for conserved syntenic regions and even local gene order within scaffolds between E. muelleri and Trichoplax adhaerens, Nematostella vectensis and the chordate Branchiostoma floridæ. Synteny conserved over hundreds of millions of years is consistent with the hypothesis that gene shuffling primarily occurs within chromosomes rather than between them, as predicted by the double cut and join-dosage sensitivity (DCJ-DS) model. The DCJ-DS model predicts that dosage sensitive genes would tend to stay on the same chromosome, although the local order may change. We find this is clearly observable in shorter chromosomes (Fig. 2d). For example, scaffold_0022 in E. muelleri matches with scaffold_5 in T. adhaerens, scaffold_3 and scaffold_16 in N. vectensis, and scaffolds Bf_V2_21 and Bf_V2_150 in B. floridæ (Fig. 2d, e). Only ten proteins from scaffold_0022 are shared by all species. Two are SL3C6V/VIAT-group transporters (Em0022g323a and Em0022g324a), one is a predicted homologue of the mitochon- drial enzyme ETFDH (Em0022g346a), and another is a predicted homologue of the splice factor A-kinase anchor protein 17A (Em0022g347a). However, 68 proteins from scaffold_0022 are found in 2 out of 3 species, generally in both T. adhaerens and B. floridæ. None of these species display co-linearity with E. muelleri in either region, indicating that while the genes match, sequential order is lost, a typical hallmark of macro synteny. However, overall, these syntenic blocks are comparable and represent ancestrally shared blocks of homologous genes, conserved from E. muelleri to T. adhaerens and B. floridæ, and thus can be inferred to represent ancient groupings conserved since the common ancestor of sponges and bilaterians. While macro-synteny has been found across genomes with high levels of contiguity, previous sponge genomes did not suggest this pattern, largely due to their comparatively fragmented assemblies (Fig. 2b).

In contrast to the conserved synteny between animal lineages, we detected no conserved syntenic regions between E. muelleri and two choano flagellates, Monosiga breviscoli and Salpingoeca rosetta (Fig. 2d, e and section 3.6 in Supplementary Note 3, Supplementary Fig. 8, Supplementary Data 4). While E. muelleri does not show clear macrosyntenic conservation with either of the two choano flagellate species do show shared synteny with each other (Supplementary Fig. 8). This disparity between the choano flagellate and metazoan gene orders suggests that gene macrosynteny has been shuffled either in the lineage leading to the ancestor of choano flagellates (Monosiga and Salpingoeca), or in the lineage leading to the last common ancestor of sponge and other animals, or, alternatively, in both of these lineages.

Pan-metazoan epigenetics. The large number of genomes now available make it clear that differences in gene regulation, as well as gene content, are responsible for the innovations seen in different animal body plans and phyla. To understand the processes underlying gene regulation in non-bilaterian metazoans, data on the three-dimensional architecture of chromatin is needed from these clades. We used HOMER and Bowtie2 to analyse Hi-C data from E. muelleri and found that as in other animals the genome is organised into topologically associating domains (TADs) as well as loops, although we did not find mammalian-like corner peaks at the edges of the predicted TADs (Fig. 3a). These TADs are slightly larger on average than those seen in Drosophila melanogaster, at 142.4 kbp, compared with ~107 kbp. As in other non-bilaterians, the E. muelleri genome lacks the CCCTC-binding factor (CTCF) but does possess a suite of non-CTCF zinc finger proteins which form a sister group to the bilaterian CTCF proteins. Besides these CTCF-like sequences, the E. muelleri genome contains both cohesin and structural maintenance of chromosome (SMC) sequences, which are highly conserved throughout eukaryotes (section 6.1 in Supplementary Note 6).

Cytosine DNA methylation is an important part of genome regulation in animals, where transcriptionally active gene bodies are methylated. However, it has recently been shown that, like vertebrates and unlike most invertebrates, sponges have highly methylated genomes in both gene bodies and intergenic regions. Whether methylation occurs in a lineage-specific manner or is driven by an exceptionally high repeat content in that species. The E. muelleri methylome thus challenges the assumption that all demosponges have hyper-methylated genomes, and suggests that the A. queenslandica pattern is a lineage-specific innovation. Whether methylation levels differ significantly in freshwater compared to marine environments has yet to be explored, especially in invertebrate taxa, and could have a bearing on this inference.
Since cytosine methylation is highly mutagenic, vertebrate and A. queenslandica genomes are highly depleted for CpG dinucleotides. Congruent with the intermediate methylation levels, we found that the genome of E. muelleri is also depleted for CpG dinucleotides, more than most invertebrates but less than in A. queenslandica (Supplementary Fig. 19). However, CpG content varies greatly across sponge genomes; for instance, S. ciliatum has higher methylation than E. muelleri, but has a relatively higher amount of CpGs. This indicates that CpG depletion is not fully coupled to methylation levels in sponges, and that retention of CpGs might obey unknown species-specific constraints.

Given that E. muelleri shows methylation levels more consistent with canonical mosaic invertebrate methylomes than with a hypermethylated genome, we then checked whether gene body methylation accumulation is dependent on gene transcription. CpGs are more commonly observed near transcriptional start sites (TSS) than in A. queenslandica, but marginally lower in absolute levels than those seen in S. ciliatum (Fig. 3c). As observed in many invertebrates, E. muelleri genes with mid-transcriptional levels show higher gene body methylation than non-expressed genes or highly expressed genes (Supplementary Fig. 19B). Promoters are strongly demethylated and repeats found within gene bodies tend to have higher methylation levels than those in intergenic regions, as seen in other invertebrates, suggesting that not all repeats are actively targeted by DNA methylation in E. muelleri. In fact, repeat methylation level positively correlates with age of the repeat, and LTR retrotransposons are more likely targeted by DNA methylation irrespective of genome position (Supplementary Fig. 19). Overall, the E. muelleri methylome shows many patterns similar to those of canonical mosaic invertebrate genomes, and may therefore provide a more appropriate comparison for future comparative epigenetics work than other existing sponge models.

Sponges show high levels of gene gain. Every sponge species we examined showed a gain of 12,000 more genes since their divergence from the most recent sister taxon or clade (Fig. 4a, section 7.1 in Supplementary Note 7, Supplementary Fig. 20). The large number of duplicates we identified in E. muelleri and other sponges is consistent with recent independent findings, and can be traced to the lineage leading to the divergence of the freshwater order Spongillida from the marine Heteroscleromorpha (Supplementary Fig. 21). This finding is robust to different placements of sponges relative to other metazoan taxa (Fig. 4a, b) and further suggests a role for duplication and gene gain in freshwater evolution.

Despite a high rate of gene gain, we observed no greater number of losses in the freshwater lineage than in other groups, and an equal number of lineage specific expansions for all taxa examined. For example, while sponges have lost 375 orthogroups compared to the last common ancestor of animals, 1340 are inferred to have been lost in the ctenophore lineage assuming that Porifera are the sister taxon to other Metazoa, or 1812 if ctenophores are assumed to be sister to other Metazoa (Fig. 4a, b, section 7.1 in Supplementary Note 7, Supplementary Fig. 21). The large numbers of genes found in sponges can therefore largely be explained by steady rates of gain in genes via duplications that are not matched by similarly high rates of gene loss.

Molecular signals of freshwater adaptation. To determine whether transitions to freshwater are accompanied by the loss of a common set of genes in independent clades, we studied shared losses in four disparate animal lineages, using pairs of species for each lineage, in which one is marine and the other is freshwater. Gene gain is also noted in section 7.4 in Supplementary Note 7, Supplementary Fig. 23, although these gains are lineage specific. Our dataset consisted of: Porifera: A. queenslandica and...
Sample text from the image:

E. muelleri; Cnidaria: Nematostella vectensis and Hydra vulgaris; Annelida: Capitella teleta and Helobdella robusta; and Mollusca: Lottia gigantea and Dreissena polymorpha. We found that, of the more than 30,000 orthogroups within which we identified losses specific to freshwater species, there were 29 shared losses in all four freshwater lineages compared to 4 shared losses in marine species (section 7.4 in Supplementary Note 7, Supplementary Data 5), while the average loss rate for any 4 taxa across our sample was 16.9 genes. We also found significantly higher ($p = 0.013$, t-test) numbers of shared orthogroup losses in three of four freshwater lineages, compared with the direct marine counterexamples of such a pattern (73, 22, 52, and 37, cf. 1, 2, 8, and 0 same-phylum marine losses, section 7.4 in Supplementary Note 7, Supplementary Data 5). Shared loss in freshwater lineages therefore seems to be an infrequent phenomenon, whereas it happens rarely across marine taxa.

The transition to fresh water, and the more recent radiation of extant species, has left signatures of positive selection in the Spongillidae24, and in E. muelleri in particular. Using multiple tests, we found 117 orthogroups to be under positive selection in E. muelleri alone, 23 of which were also under positive selection in all freshwater sponges (section 7.3 in Supplementary Note 7, Supplementary Data 6). The move to freshwater conditions must be accompanied by a diverse range of changes to membrane functionality. Several genes known to perform roles in homoeostasis and membrane function including $V$-type proton ATPase subunit B, three kinds of sorting nexin, vacuolar-sorting protein SNF8, and Multidrug and toxin extrusion protein 1, were found to be under particularly high selection pressure (Supplementary Table 11). Almost all of the E. muelleri genes in these orthogroups, 85 of 117, are differentially expressed across the process of development, underlining their importance to E. muelleri biology (Supplementary Fig. 22E). It is not uncommon for these differentially expressed genes to have multiple inparalogs. The most prolifically duplicated genes are a cytoplasmic actin and a leukotriene receptor which are both tandomly...
duplicated 6 times in the E. muelleri genome, from loci Em0009g1201 and Em0009g943 respectively. Altogether, of the 85 differentially expressed, positively selected orthogroups, 54 are single copy, and 31 possess two or more in-paralogs of the genes tested (Supplementary Data 6). This indicates that duplication of these genes is commonly associated with adaptation.

Gene duplication is known to be a means of adaptation to new environments more generally\(^3^9\), with sub-functionalisation and neo-functionalisation allowing specific changes to molecular function in response to changing conditions. We explored genes that have expanded in number in the E. muelleri genome compared to other sponges, ctenophores and chordates (here represented by B. floridiae and humans). We found that the largest clusters included genes involved in chemokine binding, and one cluster included over 50 metabotropic GABA receptors (section 7.5.1 in Supplementary Note 7). More than 120 mGABA receptors are predicted in total in the E. muelleri genome (compared to two in humans), 48% of which are on scaffolds 4, 13 and 22. Most of these are expressed, alongside other enlarged gene complements such as cortactin (52 clusters), NBAS (35) and integrin beta (36) (Supplementary Data 7).

Another gene showing high levels of duplication in E. muelleri is AQP9, a water and solute carrier known to be involved in freshwater adaptation\(^40\). Demosponges possess AQP8 (a family of aquaporins that allow the passage of mostly water, but also ammonia and urea), as well as aquaglyceroporins (AQP3,7,9 and 10), which control movement of glycerol, arsenite, and silicic acid among other compounds. However, in freshwater sponges only aquaglyceroporins are present, while many aquaporins are lost compared to outgroups (Supplementary Fig. 28). Freshwater sponges, like Hydra and many protists, use contractile vacuoles to excrete water\(^1^9\) and so it is possible that the duplication of aquaglyceroporins in freshwater sponges may have allowed some of the genes to take on new functions. For example, in mammals, AQP9 can mediate silicon influx in addition to being permeable to glycerol and urea, but not to water itself\(^4^1\). Since sponge aquaglyceroporins are more similar to AQP9 than to AQP3 or 7 (section 7.5.2 in Supplementary Note 7 and Supplementary Fig. 28) it is possible that in freshwater sponges, in particular, these genes function in silicon transport for skeletogenesis. Aquaporin-like molecules, glycerol uptake facilitator proteins (GLFPs), are found in bacteria and plants but, to date, have not been detected in animal genomes\(^4^2\). E. muelleri has nine paralogs of GLFP, with five of them located on the same scaffold (Em0019) (Supplementary Fig. 28). We hypothesise that, as in plants\(^4^2\), the presence of GLFPs in sponges came about via horizontal gene transfer.

**Gene expression during E. muelleri development.** To understand what genes are common and which are distinct from other metazoans during the development of the filter-feeding body plan, we examined differential gene expression from hatching gemmules through to the formation of a filtering sponge. The majority (32,690/39,245) of the E. muelleri gene complement was expressed at some point in the course of development. Remarkably, over 33% of the total gene models were differentially expressed (log2 (fold change) >1/−1) across the gemmule-hatching process (Fig. 5, Supplementary Note 8). The pattern of gene expression shows a typical shift that occurs during the development of animals from the breakdown of reserves stored in cells in the embryo (here thosocyte stem cells stored in over-wintering gemmules) at Stage 1. Stages 2 and 3 show activation of developmental patterning genes, genes involved in cell motility and the production of extracellular matrix. The upregulation of genes involved in structural maintenance, homeostasis, and the immune system occurs at Stage 5 (Fig. 5 and Supplementary Note 8). At Stage 1, arachidonate pathways for glycogen breakdown and fatty acid metabolism were differentially upregulated to produce the breakdown of stored reserves (Fig. 5 and Supplementary Note 8). Many of the genes involved in formation of a basement membrane and true epithelia were originally considered to be eumetazoan\(^4^3\), but we found in the E. muelleri genome, genes for type IV collagen, contactin, laminin, PAR/6, patj, perlec and nidogen that exhibit gene expression profiles consistent with their known role in development of polarised epithelia (Fig. 5 and Supplementary Fig. 31). Similarly, claudin, which may be involved in the tight seal that E. muelleri epithelia have been shown to form\(^4^4\), is expressed later in developmental time. While eukaryotic genes are expressed throughout the different developmental stages, many sponge-specific genes are expressed only as the sponge hatches and develops the aquiferous system that is common only to Porifera. Expression profiles of Wnt, TGF-β and Hedgehog signalling pathways are given in Supplementary Fig. 36 and indicate that the greatest difference in transcription occurs between hatching and the early development of the sponge specific characteristics of spicules, chambers and the aquiferous system (details, section 9.4 in Supplementary Note 9). Individual components of these pathways are expressed at discrete time points in the process of development. For example, in the Wnt pathway, three of the secreted frizzled receptor proteins (SFRP B, SFRP E, SFRP F) as well as two LRP receptors (LRP 2, LRP 4B) are upregulated during hatching from the stem cells stored in the gemmule (Stage 1), and in the TGF-β pathway, receptors are expressed in Stages 1–3, but downregulated thereafter. This data therefore provides a wealth of information for understanding and contrasting the genetic processes underpinning sponge development.

**Sensation and non-nervous signalling in sponges.** Sponges have no nervous system and yet they contract in response to a range of stimuli\(^4^5\). Exactly how contractions are coordinated is still unknown, and the potential position of ctenophores rather than sponges as sister to the rest of animals on the tree of life\(^4^6\–^4^8\) provocatively implies that sponges could have lost neurons. It has been difficult to identify a single character of neurons shared by all animal\(^4^9\), but the synapse and in particular the proteins that compose its scaffolding and chemical neurotransmitter complement, are agreed to be an important component\(^5^0\). One difficulty is that genes in the neuronal synapse also have other tissue-specific functions. One family of genes with tissue-specific functions are the Synaptosomal-associated proteins (SNAPs) of vertebrates. These proteins are members of the Soluble N-ethylmaleimide sensitive factor attachment protein RErceptor (SNARE) protein superfamily\(^5^1\), which function in membrane fusion at the cell surface. In mammalian cells SNAP-25 mediates the fusion of vesicles with the presynaptic membrane of neurons, while its parologue SNAP-23 mediates the fusion of vesicles in other regions of the cell surface\(^5^2\). The presence of genes encoding SNAP-23/25-like SNARE proteins has sometimes been inferred to indicate presence of neuron-specific protein machinery\(^5^3\). However, homologues of SNAP SNAREs are widely conserved among eukaryotes without nervous systems, such as plants\(^5^4\). This raises the question of whether particular SNAP SNARE genes found in early-branching metazoan lineages indicate an early origin of neuron-specific protein machinery.

Our phylogenetic analysis of SNAP-23/25 homologues revealed that the vertebrate neuron-specific paralogue SNAP-25 arose from a duplication that occurred in the vertebrate stem lineage, while the two SNAP-23/25-like genes found in E. muelleri arose from an independent duplication that occurred in Porifera.
choanoagellates (Supplementary Note 9). TRPA channels are some of the best characterised and are known as mechano- or chemoreceptors whereas TRPML families are largely considered to be expressed on organelles inside cells. The diversification of TRPA channels in E. muelleri and other demosponges suggests a molecular mechanism for mechanoreception as well as chemical sensation in this clade.

In the E. muelleri genome we also found a wide range of ion channels involved in signalling in eumetazoans (Supplementary Note 9), but there are conspicuous absences including voltage-gated sodium and potassium channels, epithelial sodium-activated channels (ENaCs), leak channels, and glutamate-gated ion channels (GICs). Also absent are receptors for monoamine (serotonin and dopamine) signalling, as well as key components of the biosynthesis pathways for these, as well as ionotropic glutamate receptors. While the latter are present in calcareous and homoscleromorph sponges and in non-metazoans, demosponges seem to have lost them. In contrast, we found evidence for a diversity of metabotropic glutamate receptors (mGlur), as well as a wealth of metabotropic GABA-receptors, as discussed above.

In E. muelleri therefore, as in other demosponges, there is evidence for components that allow sensation of the environment via TRP channels, among others, and non-neuronal chemical signalling via metabotropic GPCRs (e.g., receptors for glutamate and potentially GABA and/or a range of organic acids), but no evidence for more rapid electro-chemical signalling. While we find no signature for any aspect of conventional nervous tissues in
the *E. muelleri* genome, we cannot rule out the possibility that the phylum Porifera as a whole, or individual lineages within it (including *E. muelleri*), have lost these neuron-related components.

**Host–microbe associations in *E. muelleri***. Most animals possess diverse symbiotic microbial consortia, which provide their hosts with metabolic advantages and new functions, and sponges are no different. The release of the genome of *A. queenslandica* opened a window into the study of the mechanisms of sponge–microorganism interactions. To unravel the recognition mechanisms developed by host and microbes to facilitate symbioses, high quality genomes (and more genomic resources in general) are fundamental.

The genome of *E. muelleri* offers a model that allows exploration of eukaryotic patterns of microbial recognition in unique environments. We studied the microbiome of 11 different specimens of *E. muelleri* collected from six locations across 6500 km in the Northern hemisphere, and found that this species contained between 865 and 4172 unique amplicon sequence variants (ASVs) (Supplementary Note 10, Supplementary Data 9). The microbiome of *E. muelleri* has a level of diversity comparable to that of the most diverse marine demosponges. The microbiome of all specimens of *E. muelleri* is largely dominated by Proteobacteria and Bacteroidetes, as in other demosponges (Supplementary Note 10). However, like other freshwater sponges, *E. muelleri* possesses a large fraction of the order Betaproteobacteriales, absent in marine sponges, which are traditionally associated with Gammaproteobacteria, a difference which is likely due to the differing pH and nutrients found in the two environments. Surprisingly, even though the entire genome of an unknown species of *Flavobacterium* was recovered during the sponge sequencing (Supplementary Note 4), Flavobacteriales were not especially abundant in the other *E. muelleri* samples, only reaching 16% relative abundance in adult tissue from UK samples (Supplementary Note 10).

Overall, differences in microbiome content were determined by geographic location, as has been found in marine sponges. For example, only the samples collected from the Sooke Reservoir had a high abundance of Firmicutes and Campylobacteria. Likewise, only those samples collected from Maine had a moderate abundance of Cyanobacteria. Despite the distance separating samples, and therefore potential different ecologies of the collection sites, we found that four ASVs were shared among all samples, yet with different percentages ranging from <1% to >20% in different samples. These four ASVs were assigned to Burkholderiaceae (order Betaproteobacteriales) and Ferruginibacter (order Chitinophagales), and one was an unclassified bacterium (Supplementary Note 10). Whether these ASVs are fundamental for the metabolic function of *E. muelleri*, or whether they are simply cosmopolitan bacteria transported by the wind or on animals, and taken up by all sponges in lakes and rivers, is still to be determined. These findings and resources open the door to studies of species-specific patterns of host–microbe association at a broad scale.

**Conclusions**. The high quality of the *E. muelleri* genome provides a new basis for comparative studies of animal evolution. To date we have lacked a chromosomal-quality poriferan genome assembly, and with this in hand for an experimentally tractable organism, comparative studies of a variety of ancestral characters, including longer-range gene regulation and genomic architecture, become possible.

Given their apparent anatomical simplicity, it can be surprising to some researchers that sponges have nearly twice the gene complement of other animals, but the high quality of this genome confirms that this is not an artefact of previous genome assemblies, and suggests that gene duplication and adaptation to novel environments are responsible for the high gene counts. Sponges possess complex filtering behaviours, integrate with an extensive network of microbes, and have an extensive defence system. As only approximately half of the genes found in sponges can be firmly identified, it is clear that there remains a huge amount of hidden biology yet to be understood in sponges, just as in other non-bilaterians. The robustness of the *E. muelleri* genome and model is an excellent tool for performing this work. It also opens the door to comparative analysis of the genomic changes required for the challenging process of adaptation to fresh water, and to finding out whether these are shared convergently in disparate phyla. Complemented by additional RNAseq, methylation data, and the analysis of symbiont content, the *E. muelleri* genome offers an important new opportunity for exploring the molecular toolkit, from protein coding to gene regulation, that underpinned the early evolution of animals and their diverse, complex, and successful traits.

**Methods**

**Sequencing and assembly**. Tissue used for DNA sequencing was derived from a single clone collected as overwintering cysts (gemmae) from the Sooke Reservoir, at the head tank of the city of Victoria, British Columbia drinking water system. A voucher specimen is deposited with the Royal British Columbia Museum (RBCM519-00160-001) (Supplementary Notes 1 and 2). Tissue from a single clone hatched and grown under sterile conditions was flash frozen and stored at −80 °C. DNA isolation and sequencing was carried out by Dovetail Genomics (Scotts Valley, CA, USA) using PacBio sequencing for de novo assembly with Arrow (genomiconcensus package, PacBio tools) followed by preparation of Chicago and HiC libraries that were sequenced on Illumina platforms, and subsequent assembly using HiRISE (Supplementary Note 2). Genome assembly metrics were determined using a range of tools, and further details of all methods used are available in Supplementary Note 2.

**Genome annotation**. Gene models were predicted using AUGUSTUS 3.3.2 annotation software (http://bioinf.uni-griffiswald.de/augustus/) with previously published RNAseq datasets used for training. As the basis for gene prediction, the non-masked genome was used, to avoid artefacts, missed exons or missing gene portions caused by masked areas of the genome. The BUSCO v2.3 set was used in genome mode to determine gene recovery metrics. RepeatModeler 2.0 and RepeatMasker 4.1.0 were used sequentially to predict repetitive content within the genome as described in Supplementary Note 3. Contamination and bacterial content was excluded by BLAST 2.10.0 against a range of well-annotated databases. Syntenic relationships were assessed using reciprocal blasts and custom Python 3.7 scripts (scaphid_symenv.py, see bitbucket repository http://bitbucket.org/EphydatiaGenome/output Genome/ or available in Supplementary Data 2). Taxonomy assessment of identified sambiont sequence was performed in MGAlign2 and other software, as detailed in Supplementary Note 4.

Automated annotation of gene sequences was performed using DIAMOND 0.9.31 BLASTX against the nr and Swiss-Prot databases followed by functional annotation. Full details are described in Supplementary Note 5. Methylation studies were performed using the Methylation-seq protocol and as described in Supplementary Note 6. TADS and loops were identified using HOMER v4.115. Orthogroup based analyses were performed primarily using Orthofinder26, with FROG v1.0.6, MAFFT 7.580, and DIAMOND BLAST options (described in detail Supplementary Note 7). Selection tests were performed according to the methodology put forward by Santagata (https://github.com/Santagata/Select_Test) and Kenny et al.24 and detailed in Supplementary Note 7.3.

**Developmental gene expression**. RNAseq was performed by LC Sciences (Houston, TX). HISAT 2.0.67 was used to map RNAseq reads to the reference *E. muelleri* genome. edgeR v3.14.088 was used to estimate the expression levels of all transcripts across all replicate samples. Full details are provided in Supplementary Note 8.

Gene family content was assessed using targeted manual BLAST, with HMMER v3.069 used when necessary to test absence. Reciprocal BLAST was used to ensure assignment of identity, with the identity of key gene families assessed using phylogenetic inference as shown in Supplementary Note 9. Holobiont content from a number of *E. muelleri* samples was assessed with Mothur v1.41.3 and an adaptation of MiSeq SOP protocol99 as fully described in Supplementary Note 10.
References
1. Richter, D. J. & King, N. The genomic and cellular foundations of animal origins. Ann. Rev. Genet. 47, 509–537 (2013).
2. King, N. & Rokas, A. Embracing uncertainty in reconstructing early animal evolution. Curr. Biol. 27, R1081–R1088 (2017).
3. Telford, M. J., Moroz, L. L. & Halanych, K. M. A sisterly dispute. Nature 529, 286–287 (2016).
4. Srivastava, M. et al. The Amphimedon queenslandica genome and the evolution of animal complexity. Nature 466, 720–727 (2010).
5. Fernandez-Valverde, S. L., Calcino, A. D. & Degnan, B. M. Deep developmental transcriptome sequencing uncovers numerous new genes and enhances gene annotation in the sponge Amphimedon queenslandica. BMC Genomics 16, 387–387 (2015).
6. Fernandez-Valverde, S. L. & Degnan, B. M. Bilaterian-like promoters in the highly compact Amphimedon queenslandica genome. Sci. Rep. 6, 22496 (2016).
7. Leininger, S. et al. Developmental gene expression provides clues to relationships between sponge and eumetazoan body plans. Nat. Commun. 5, 3905 (2014).
8. Francis, W. R. et al. The genome of the contractile demosponge Tethya wilhelma and the evolution of metazoan neuronal signalling pathways. Preprint at BioRxiv https://doi.org/10.1101/120998 (2017).
9. de Mendoza, A. et al. Convergent evolution of a vertebrate-like myoglobin in a marine sponge. Nat. Ecol. Evol. 3, 1464–1473 (2019).
10. Larroux, C. et al. Genesis and expansion of metazoan transcription factor gene classes. Mol. Biol. Evol. 25, 980–996 (2008).
11. Van Soest, R. W. M. et al. Global diversity of sponges (Porifera). PLoS ONE 7, e35105 (2012).
12. Dohrmann, M. & Wörheide, G. Dating early animal evolution using phylogenomic data. Sci. Rep. 7, 3599–3599 (2017).
13. Adamska, M., Degnan, B. M., Green, K. & Zawfank, C. What sponges can tell us about the evolution of developmental processes. Zoology 114, 1–10 (2011).
14. Leys, S. P. & Hill, A. The physiology and molecular biology of sponge tissues. Adv. Mar. Biol. 62, 1–56 (2012).
15. Reisingw, H. M. The aquatic systems of three marine Demospongiae. J. Morphol. 145, 493–502 (1975).
16. Leys, S. P., Mackie, G. O. & Reisingw, H. The biology of glass sponges. Adv. Mar. Biol. 52, 1–145 (2007).
17. Vacelet, J. & Boury-Esnault, N. Carnivorous sponges. Nature 373, 333–335 (1994).
18. Schuster, A. et al. Divergence times in demosponges (Porifera); first insights from new mitogenomes and the inclusion of fossils in a birth-death clock model. BMC Evol. Biol. 18, 114 (2018).
19. Brauer, E. B. Osmoregulation in the freshwater sponge, Spongilla lacustris. J. Exp. Zool. 192, 181–192 (1973).
20. Barbeau, M. A., Reisingw, H. M. & Rath, L. C. Hatching of freshwater sponge gemmules after low temperature exposure: Ephydatia malloti (Porifera: Spongillidae). J. Therm. Biol. 14, 225–231 (1989).
21. Manconi, R. & Pronzato, R. in Porifera Research: Biodiversity, Innovation and Sustainability. Vol. 28 (eds Custodio, M. R. et al.) 61–77 (Livros de Museu Nacional, Lisboa, 2007).
22. Manconi, R. & Pronzato, R. in Freshwater Animal Diversity Assessment (eds Balian, E. V. et al.) 27–33 (Springer, Dordrecht, 2008).
23. Späcke, J. Wastewater plant’s success spawns salmon and sponges Water Wastes Digest. http://www.wwmdmag.com/wastewater/wastewater-plants-across-salmon-sponges/ (2000).
24. Kenny, N. J., Plese, B., Riesgo, A. & Itochik, V. B. Freshwater adaptation at the molecular scale in the unique sponges of Lake Baikal. Mol. Biol. Evol. 36, 2462–2480 (2019).
25. Mukai, H. Further studies on growth and sex differentiation in four species of freshwater sponges. Sci. Rep. Fac. Edu Gunma Univ. 39, 41–56 (1999).
26. Elliott, G. & Leys, S. Coordinated behaviour in a freshwater sponge: orchestrated contractions effectively expel water from the aqueous system. J. Exp. Biol. 210, 3736–3748 (2007).
27. Rivera, A. et al. The evolution and function of the Pax/Six regulatory network in sponges. Evol. Dev. 15, 189–196 (2013).
28. Francis, W. R. & Wörheide, G. Similar ratios of introns to intergenic sequence across animal genomes. Genome Biol. Evol. 9, 1582–1598 (2017).
29. Simakov, O. et al. Deeply conserved synteny resolves early events in vertebrate evolution. Nat. Ecol. Evol. 4, 820–830 (2020).
30. Fernández, R. & Gabaldón, T. Gene gain and loss across the metazoan tree of life. Nat. Ecol. Evol. 4, 524–533 (2020).
31. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol. Cell 48, 471–484 (2012).
32. Gaiti, F. et al. Landscape of histone modifications in a sponge reveals the origin of animal cis-regulatory complexity. elife 6, e22194 (2017).
33. Zemach, A. & Zilberman, D. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. J. Exp. Biol. 218, R780–R785 (2015).
34. Gatzmann, F. et al. The methylene of the marbled crayfish links gene body methylation to stable expression of poorly accessible genes. Epigenet. Chromatin 11, 57–57 (2018).
35. Kondrashov, F. A. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc. Biol. Sci. 279, 5048–5057 (2012).
36. Heymann, J. B. & Engel, A. Aquaporin phylogeny, structure, and physiology of water channels. Physiology 14, 187–199 (1999).
37. King, L. S., Kozono, D. & Agre, P. From structure to disease: the evolving tale of aquaporin biology. Nat. Rev. Mol. Cell Biol. 5, 687–698 (2004).
38. Gustavsson, S., Lebrun, A.-S., Nordén, K., Chaumont, F. & Johanson, U. A novel plant major intrinsic protein in Physcomitrella patens most similar to bacterial glycerol channels. Plant Physiol. 139, 287–295 (2005).
39. Fahey, B. & Degnan, B. M. Origin of animal epithelia: insights from the sponge genome. EvoDevo 12, 601–617 (2010).
40. Adams, E., Goss, G. & Leys, S. Freshwater sponges have functional, sealing epithelia with high transepithelial resistance and negative transepithelial potential. PLoS ONE 5, e13040 (2010).
41. Leys, S. P. Elements of a ‘nervous system’ in sponges. J. Exp. Biol. 218, 581–591 (2015).
42. Dunn, C. W. et al. Broad phylogencic sampling improves resolution of the animal tree of life. Nature 452, 745–749 (2008).
43. Ryan, J. F. et al. The genome of the ctenophore Mnemiopsis leidyi and its implications for cell evolution. Science 342, 124592 (2013).
44. Moroz, L. L. et al. The ctenophore genome and the evolutionary origins of neural systems. Nature 510, 109–114 (2014).
45. Anderson, P. A. V. On the origins of those most transformative of biological systems—the nervous system. J. Exp. Biol. 218, 504–505 (2015).
46. Sakaraya, O. et al. A post-synaptic scaffold at the origin of the animal kingdom. PLoS ONE 2, e506 (2007).
47. Jahn, R. & Scheller, R. H. SNAREs—engines for membrane fusion. Nat. Rev. Mol. Cell Biol. 7, 631–643 (2006).
48. Delgado-Martínez, I., Nehring, R. B. & Sørensen, J. B. Differential abilities of agellates lead the way. Nature 452, 514 (2008).
49. Burkhardt, P. The origin and evolution of synaptic proteins—choanoligaselets heal the way, J. Exp. Biol. 218, 506–514 (2015).
54. Sanderfoot, A. Increases in the number of SNARE genes parallels the rise of multicellularity among the green plants. Plant Physiol. 144, 6–17 (2007).

55. Ludeman, D., Farrar, N., Riesgo, A., Paps, J. & Leys, S. P. Evolutionary origins of sensation in metazoa: functional evidence for a new sensory organ in sponges. BMC Evol. Biol. 14, 3 (2014).

56. Peng, G., Shi, X. & Kadowaki, T. Evolution of TRP channels inferred by their classification in diverse animal species. Mol. Phy Evol. 84, 145–157 (2015).

57. Hentschel, U., Piel, J., Degnan, S. M. & Taylor, M. W. Genomic insights into the marine sponge microbiome. Nat. Rev. Micro. 10, 641–654 (2012).

58. Thomas, T. et al. Diversity, structure and convergent evolution of the global sponge microbiome. Nat. Commun. 7, 11870–11870 (2016).

59. Laport, M. S., Pinheiro, U. & Rachid, C. T. C. C. Freshwater sponge Tubella variabilis presents richer microbiota than marine sponge species. Front. Microbiol. 10, 2799 (2019).

60. Dunn, C. W., Leys, S. P. & Haddock, S. H. D. The hidden biology of sponges and ctenophores. Curr. Biol. 30, 282–291 (2015).

61. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212 (2015).

62. Rodríguez-R, L. M. et al. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and bacteria at the whole genome level. Nucleic Acids Res. 46, W282–W288 (2018).

63. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60 (2015).

64. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157 (2015).

65. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360 (2015).

66. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2009).

67. Wheeler, E. J. & Eddy, S. R. nhmmer: DNA homology search with profile HMMs. Bioinformatics 29, 2487–2489 (2013).

68. Edgar, R. C. Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. PeerJ 5, e3889 (2017).

Acknowledgements

Genome and RNA sequencing for this project was funded by a Natural Sciences and Engineering Research Council (Canada) Discovery Grant (RGPIN-2016-05446) to S.P.L. This work was supported by the EU Horizon 2020 MSCA ADAPTOMICS (grant agreement: H759097) to N.J.K. and Deepyom (grant agreement: 796011) to C.D.V. The selection test methods applied in this study were facilitated using best practices and scripts provided as part of a Next Generation Sequencing–based workshop sponsored by the National Science Foundation (Award # 1744877 to S. Santagata). G.W., K.J., R.E.R.V. acknowledge funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 768480 (ITN IGNITE). A.L.H. and S.C. acknowledge funding from the National Science Foundation (Award #1555440). I.F.R. acknowledges funding from the National Science Foundation (Award No 1542597). A.R. acknowledges funding from the EU SponGES project (European Union’s Horizon 2020 research and innovation programme, grant agreement: 679849). A.R., G.W., W.R.F. acknowledge funding from VILLUM FONDEN (Grant No. 16518). L.B.-C. acknowledges funding from the DFG, grant number: JE 777/3-1. We thank Karen Evans for sending amplicon sequencing samples, Glen Elliott, Angela Bentley (SEM) and Pamela Windsor-Reid (silicatein FISH) for help in providing images for Fig. 5 and Supplementary Fig. 40, and Isabelle Oliver for help in constructing Supplementary Fig. 35. Ilan Dornich drew the sponge illustrations in Figs. 1 and 5. We thank Dr Ferdinand Marlétaz and Dr Jordi Paps for their comments and guidance with this manuscript, and the members of the Leys and Riesgo labs for their myriad help and support.

Author contributions

S.P.L. conceived of the project, performed gene analysis, wrote and managed the overall direction of this manuscript. N.J.K., W.R.F. and A.R. performed bioinformatic analyses, wrote the manuscript and aided in project coordination. A.L.H. led work identifying genes, and particularly analysis of Wnt family content, wrote the manuscript and aided in project coordination. G.W. helped write the manuscript and aided in project coordination. S.C. aided with many aspects of gene identification. R.E.R.V. contributed specialist analyses on genome architecture and symbiont content. K.J. provided analyses of unique genes in E. muelleri and cluster expansions. A.D.M. and R.L. contributed data and analysis of methylation in E. muelleri and other species. J.F.R. contributed to studies of loss, and provided specialist bioinformatic assistance. L.B.-C. and L.D.B. provided data and analysis on T.R.P. and SNARE genes. I.G. aided in sample collection, storage and processing. M.R. aided with analysis of Hi-C contact data and inference of genome architecture. C.D.V. performed amplicon sequencing, analysis and curation of results. All authors contributed to and approved the manuscript before submission.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.J.K. or S.P.L.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

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

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