





RESEARCH ARTICLE

Open Access



Emergence of distinct syntenic density regimes is associated with early metazoan genomic transitions

Nicolas S. M. Robert^{1*} , Fatih Sarigol¹ , Bob Zimmermann¹ , Axel Meyer² , Christian R. Voolstra²  and Oleg Simakov^{1*} 

Abstract

Background: Animal genomes are strikingly conserved in terms of local gene order (microsynteny). While some of these microsynteny have been shown to be coregulated or to form gene regulatory blocks, the diversity of their genomic and regulatory properties across the metazoan tree of life remains largely unknown.

Results: Our comparative analyses of 49 animal genomes reveal that the largest gains of synteny occurred in the last common ancestor of bilaterians and cnidarians and in that of bilaterians. Depending on their node of emergence, we further show that novel syntenic blocks are characterized by distinct functional compositions (Gene Ontology terms enrichment) and gene density properties, such as high, average and low gene density regimes. This is particularly pronounced among bilaterian novel microsynteny, most of which fall into high gene density regime associated with higher gene coexpression levels. Conversely, a majority of vertebrate novel microsynteny display a low gene density regime associated with lower gene coexpression levels.

Conclusions: Our study provides first evidence for evolutionary transitions between different modes of microsyntenic block regulation that coincide with key events of metazoan evolution. Moreover, the microsyntenic profiling strategy and interactive online application (Syntenic Density Browser, available at: <http://synteny.csb.univie.ac.at/>) we present here can be used to explore regulatory properties of microsyntenic blocks and predict their coexpression in a wide-range of animal genomes.

Background

Local gene order has been conserved across animal phyla over vast evolutionary time spans, and is referred to as microsynteny [1–5]. However, little is known about the loss of ancestrally inherited microsynteny and the emergence of novel microsynteny, due to genome rearrangements, gene gains and/or gene losses [1, 6–8]. Determining the node of microsynteny loss or emergence can provide insights into the evolution of animal

genome architecture and the extent of microsynteny conservation.

It still remains unclear whether microsynteny is conserved due to functional constraints (e.g., *cis*-regulatory constraints, topological organization) [9–11], or if it is simply a result of low recombination rates (i.e. without functional significance for gene regulation). Only few studies have addressed the possible functional roles of microsyntenic blocks across the whole genome. There is increasing evidence that some evolutionarily conserved pairs of adjacent genes are maintained as gene regulatory blocks (GRBs) because of *cis*-regulatory constraints, i.e., the regulatory regions of a target gene are located within a so-called bystander, an unrelated neighboring gene [6,

*Correspondence: nicolas.robert@univie.ac.at; oleg.simakov@univie.ac.at

¹ Department of Neurosciences and Developmental Biology, University of Vienna, Althanstrasse 14, 1090 Wien, Austria

Full list of author information is available at the end of the article



9, 10]. In human GRBs, the expression of bystander and target genes does not correlate [12]. However, genes within microsyntenic blocks of at least three genes display a higher than expected coexpression in invertebrate genomes [13]. The regulatory constraints on the expression of conserved gene pairs are well characterized in some species [6, 12], whereas those on microsyntenic blocks comprising three or more genes are largely lacking, due to both missing genomic information to identify such regions as well as missing functional genomic data.

Gene density has been suggested as a proxy for inferring the level of coexpression of genes, as a positive correlation between gene proximity and gene coexpression has also been reported [14, 15]. Previous studies have shown that microsyntenic blocks can possess a gene density diverging from the genome average in some species. For instance, a genome-wide study of vertebrate amniotes showed that the conserved syntenic regions exhibited a lower gene density than the rest of the genome [16]. In contrast, the *Lrk* gene loci were shown to have a high gene density in wheat, barley, maize, and rice [17].

In this context, our objective was to conduct an exhaustive search among available animal genomes to profile the emergence of conserved microsynteny across the animal tree of life and to investigate their genomic properties. We used comparative genomics approaches to determine the retention of gene density across orthologous microsyntenic blocks and, in conjunction with available developmental RNAseq data in a few phylogenetically dispersed animals, to study the impact of microsyntenic gene density on gene co-expression. Our results provide new insights into the evolution and functional significance of conserved microsyntenic genome architecture.

Results & discussion

Largest gain of microsynteny in the bilaterian and planulozoan ancestors

The increased taxonomic coverage of sequenced genomes allowed us to profile syntenic gains and losses across the metazoan tree. According to previous methods for synteny detection across distantly related animals [1, 2, 13], we define a microsyntenic block as a unit of three or more orthologous genes, each separated by up to five intervening genes (Supplementary Fig. 2A). As local order has been scrambled across the vast evolutionary distances [1, 18], we do not require the orthologs to exhibit conserved collinearity. Our focus on blocks of three or more orthologous genes allows for a dynamic capture of longer microsyntenic stretches and the threshold of five intervening genes minimizes the proportion of false positives [1].

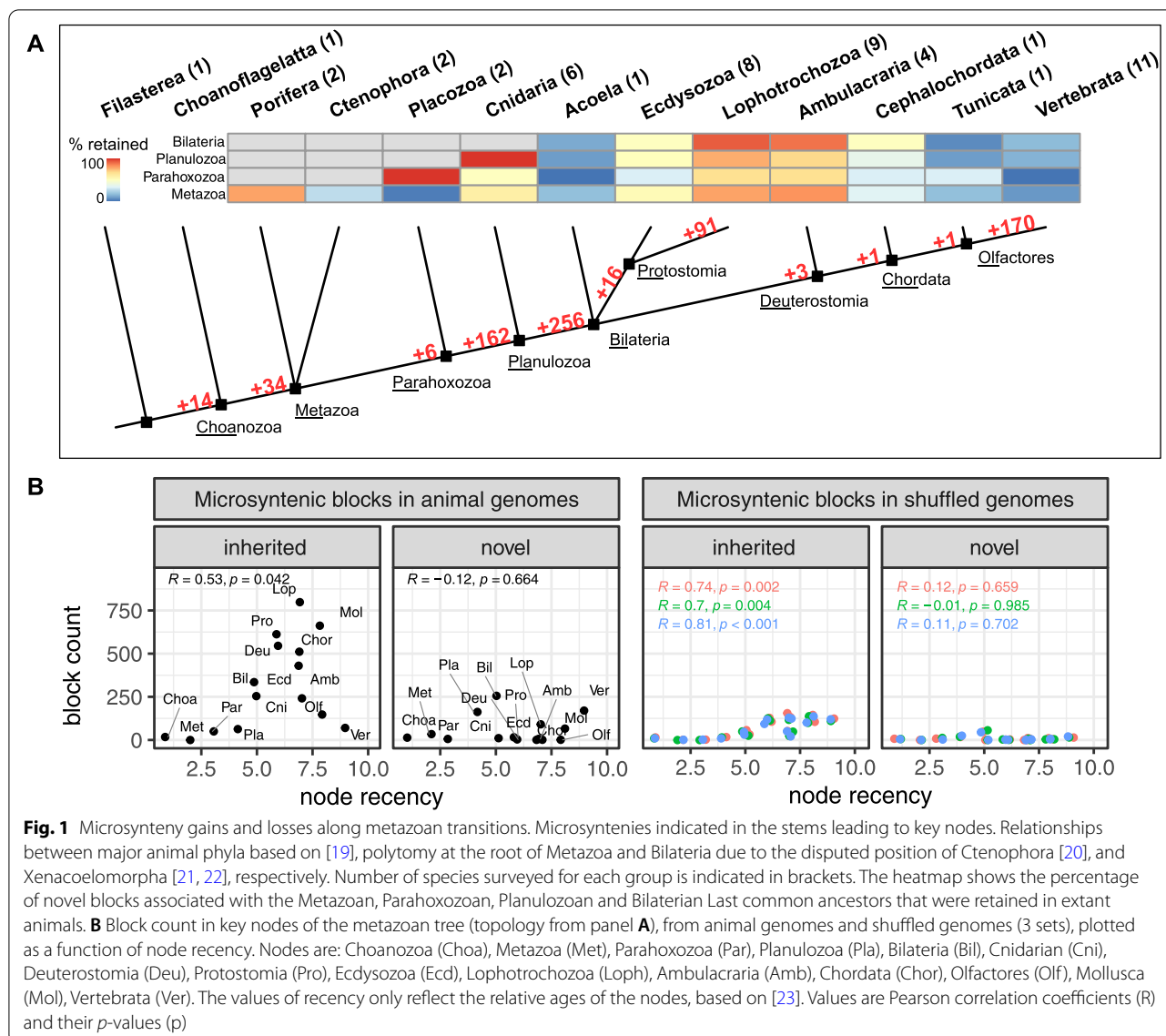
Using the genomes of 49 metazoan species from 18 phyla (Supplementary Fig. 1) we reconstructed the

ancestral complement of microsyntenic blocks throughout metazoan evolution (Fig. 1A). For every investigated node, we distinguished between inherited multi-species blocks (found in both the ingroup and the outgroup) and novel multi-species blocks (blocks found in the ingroup but not in the outgroup, see Supplementary Fig. 2B, and Methods). To assess false positive detection, we inferred microsyntenic blocks in three sets of shuffled genomes, i.e. genomes where the relative positions of genes have been randomly reshuffled (Methods).

We infer the largest gains of microsyntenic blocks to have occurred in the bilaterian last common ancestor (BLCA, 256 new blocks, only 45, 48 or 56 blocks found in shuffled genomes), and the planulozoan last common ancestor (PLCA, 162 new blocks, only 17, 19 or 25 in shuffled genomes) (Fig. 1B). In contrast, only 34 new blocks were found in the metazoan last common ancestor (MLCA, 0 in any shuffled genome). Thus, a substantial proportion (43%) of the microsyntenic blocks found in the BLCA emerged after the cnidarian-bilaterian split. Given the disputed position of the ctenophore and xenacoelomorpha phyla [21, 22, 24, 25], we calculated the amount of microsyntenic blocks novelties under various phylogenetic hypotheses. Regardless of the hypothesis, we obtained similar results regarding the counts of MLCA, PLCA and BLCA novel blocks (Supplementary Fig. 3). Accordingly, for every subsequent analysis, we used the set of blocks inferred using the multifurcating tree in Supplementary Fig. 1.

We next investigated the retention of microsyntenic blocks in extant species, defining a block to be lost if no species of a given clade possessed it in our analysis. The BLCA novelties have been largely retained among ambulacrarians and lophotrochozoans (Fig. 1A). Conversely, vertebrates lost 84% of the BLCA blocks, and ecdysozoans lost 50%. We observed medium to high percentages of lost BLCA novel blocks in acoel, tunicate, and cephalochordate (88, 95, and 51%, respectively). Since these loss rates are impacted by species sampling, the accuracy of these estimates will increase as more genomes become available for those clades. Nevertheless, the extensive loss of ancestral synteny in vertebrates can be explained by the whole genome duplications that were followed by paralog loss [26–29]. In ecdysozoans, this loss of synteny is explained by accelerated evolutionary rates involving a large number of chromosome fusions and rearrangements [30, 31] and extensive gene loss [32, 33].

It is expected that the number of genomic rearrangements between two taxa is proportional to the evolutionary time since common ancestry. In order to confirm that our estimations of the number of novel blocks were not biased by the timing of divergence, we correlated the inferred microsynteny block counts to phylogenetic



node recency (a higher “recency” value means that the last common ancestor (LCA) is younger, [Methods](#)). In observed blocks, while the number of inherited blocks positively correlated with node recency (Pearson correlation coefficient 0.53, $p < 0.05$), the number of novel blocks did not (Fig. 1B), supporting the observation of extensive microsyntenic gains in the PLCA and BLCA. The positive correlation that we see between node recency and the associated number of inherited blocks in shuffled genomes can be explained by the increasing size of the outgroups, a larger outgroup leading to more blocks detected by chance. In most of the ancestors of the different metazoan lineages, the number of inherited blocks is several times lower in shuffled genomes. In contrast, the vertebrate LCA inherited only 70 blocks, less than the

110, 120 and 124 expected to occur by chance (Fig. 1B). The loss of microsyntenic blocks that followed the whole genome duplication events in vertebrates affected not only the BLCA novel microsynteny, but even microsyntenic blocks inherited from older nodes, as exemplified by their loss of 91 and 85% of the MLCA and PLCA novelties, respectively (Fig. 1A).

Syntenic gains at various points in evolution were associated with different molecular functions. Among the 212 Gene Ontology (GO) terms enriched in the genes within the BLCA novel synteny (Supplementary Table 1), some of the most significant ones were cell communication, signaling, and the establishment or maintenance of chromatin architecture ([Methods](#), Supplementary Dataset 2). This contrasts with the 31 GO terms enriched in MLCA

novel syntenies (Supplementary Table 3), mostly associated with the emergence of multicellularity, for example protein complex involved in cell adhesion, immune system process, and heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules.

Taken together, our results demonstrate that the early evolutionary transitions towards bilaterally symmetrical animals [34] were associated with the emergence or higher retention of many microsyntenic blocks, as reflected in the substantial increase already in the planulozoan node (Fig. 1A). This is in strong contrast to the preceding phylogenetic transition and the emergence of multicellularity in the metazoan stem [35] which was associated with novel gene emergence [4, 36–38]. This suggests that different mechanisms of genomic innovation may underlie early animal evolutionary transitions.

When investigating the functional significance of gene order, it is important to also consider the three-dimensional structure of the chromatin. Topologically associating domains or TADs, for example, are regions of the genome that show a high degree of internal physical interaction, but little to no physical interaction with their neighboring regions [39]. To date, much of the relationship between microsynteny and TADs remain unexplored, due to the lack of high-resolution HiC data for many invertebrate species, and only few syntenic regions have been described [11, 40]. Previous studies determined that bilaterians TAD formation is associated with CCCTC-binding factor (CTCF) proteins [39, 41]. The fact that CTCF protein could only be detected among bilaterians [42] suggests that the spatial organization of bilaterian chromatin differs from that of non-bilaterians. We hypothesize that distinct constraints on genome topology should be reflected in the syntenic signal. This is consistent with our observation that the evolutionary split between Cnidaria and Bilateria is associated with major changes of both the microsyntenic complement (Fig. 1A) and its genomic properties (Fig. 2). Accordingly, CTCF-dependent chromatin looping could be one of the forces underlying the maintenance of bilaterian microsyntenies [43], whereas the maintenance of PLCA novel blocks in

extant cnidarians (devoid of CTCF) implies other constraints that predate the proposed emergence of CTCF.

Distinct gene density properties of microsyntenic blocks

Next, we sought to describe the genomic properties of MLCA, PLCA, and BLCA novel microsyntenies across metazoans. Not only has the proximity of two genes been suggested to facilitate their coexpression [14, 15], but the topological organization of genomes also correlates with differences in local gene density in some species, e.g., in *Drosophila* the density of TAD interior and boundaries were shown to exhibit low and high gene density, respectively [44]. Therefore, gene density within microsyntenies may comprise a proxy measure for regulatory or topological properties. However, the comparison of block density across species is hampered by the differences in gene number and genome size, two parameters that are not tightly correlated in eukaryotes [45]. In order to account for differences in density across multiple genomes, we normalized block density (defined as the total number of protein-coding genes within a microsyntenic block, including intervening genes, divided by the block length) by the whole-genome density (total number of coding genes divided by the assembly size) (see Methods). This normalization was used because the reciprocal of gene density increases linearly with the assembly size (Supplementary Fig. 4). We term this value the “normalized gene density” within microsyntenic blocks. In order to determine whether normalized gene density in microsyntenic blocks differed from the rest of the genome, we compared it to that of randomly sampled blocks, i.e. regions of the genomes sampled based on the parameters of the observed blocks (number of syntenic genes, number of intervening genes; see Methods). As errors in coding exon predictions may affect our gene density estimates, we complemented this measure of the “normalized gene density” with the median intergenic distance between consecutive genes of each block (Supplementary Fig. 7).

Our analysis showed that syntenic blocks can have higher or lower normalized gene density and median intergenic distances than those of random blocks. We

(See figure on next page.)

Fig. 2 Gene density of microsyntenic novelties associated with key metazoan transitions and their scaling across bilaterians. **A** Number of microsyntenic blocks (including duplicates/split ones) inherited from key nodes (columns) found in extant metazoan taxa (rows). Colors note whether median normalized gene density of the block is higher in observed or randomly sampled blocks (see legend). Vivid colors note a statistically significant difference ($p < 0.05$, Wilcoxon rank-sum test). Left, non-paralogous blocks (less than 40% of syntenic genes belong to the same orthogroup); right, paralogous blocks (more than 40% of genes belong to the same orthogroup). Distributions of normalized gene density can be found in Supplementary Fig. 5. **B** Heatmaps showing the Spearman correlation coefficient of the relative deviation to random density of orthologous blocks, for every taxon pair sharing at least 10 blocks. Missing data is shown in grey. **C** Scatterplots displaying the relative change of normalized gene density (Methods) of blocks shared between lophotrochozoans and ambulacrarians, vertebrates or ecdysozoans blocks. For the novel MLCA, PLCA and BLCA blocks inherited by the aforementioned taxa, the correlation coefficients and p-values are indicated if the number of blocks shared is more than 10. The scatterplots for every possible taxon pair of our sample can be found in Supplementary Fig. 6

A

	Blocks composed of non-paralogous genes						Blocks composed of paralogous genes					
	Meta	Para	Planu	Bila	Vert	Loph	Meta	Para	Planu	Bila	Vert	Loph
Vertebrate	1	0	24	49	674		2	0	26	31	17	
Tunicate	3	0	2	5			5	2	14	8		
Cephalochordate	4	1	40	101			23	5	28	41		
Ambulacrarian	26	2	164	410			70	5	122	159		
Lophotrochozoan	28	2	169	492		243	89	15	245	196		116
Ecdysozoan	15	0	81	119			60	4	114	99		
Acoel	1	0	2	6			5	0	16	29		
Cnidarian	18	2	277				40	2	186			
Placozoa	0	1					3	14				
Ctenophore	5						15					
Poriferan	8						63					

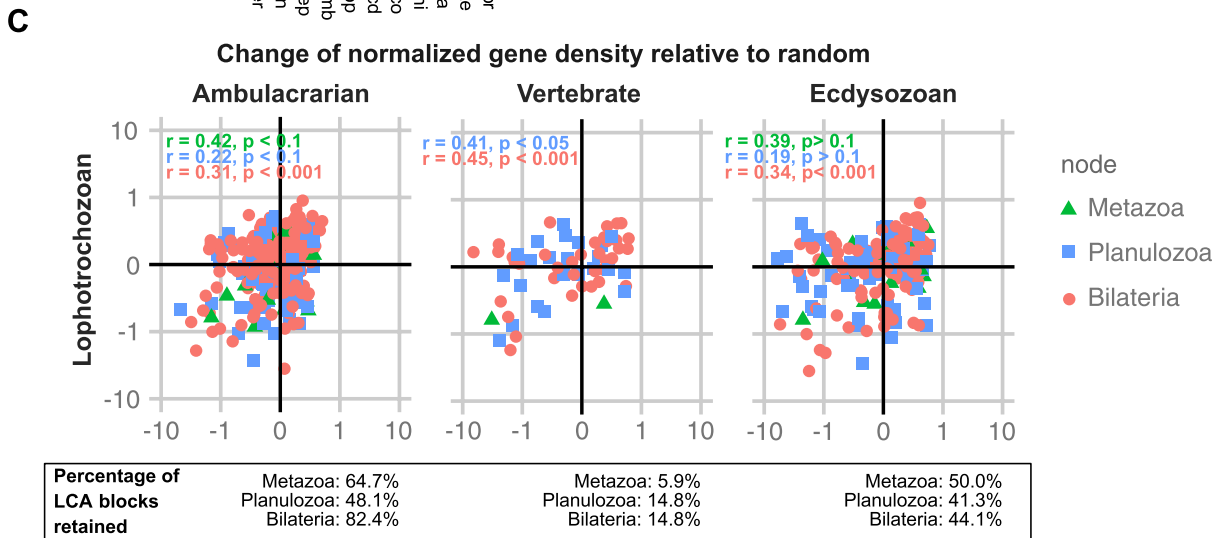
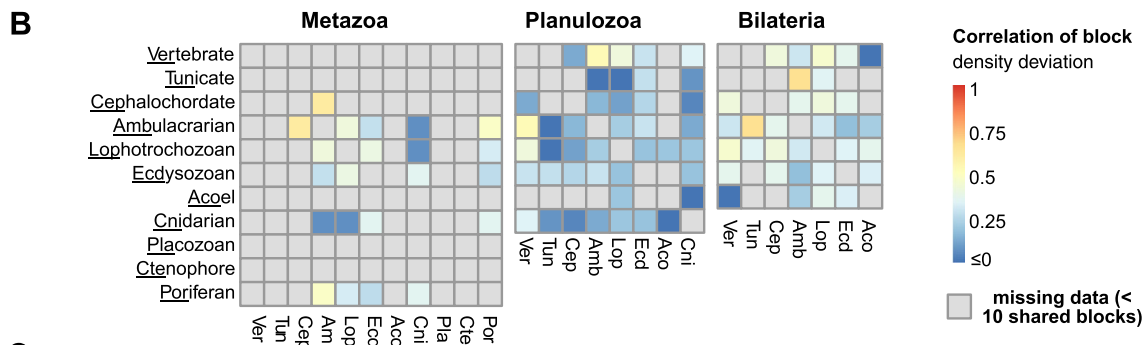
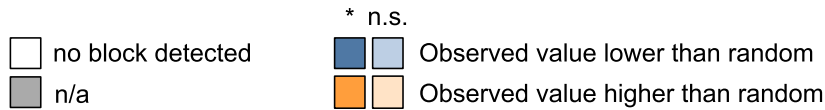


Fig. 2 (See legend on previous page.)

respectively term the blocks as following a high or a low density regime.

A major force that affects gene density regimes is tandem duplication, which can create gene-dense regions [46, 47]. To account for this, we classified microsynteny into paralogous (more than 40% of genes are paralogs) and non-paralogous (40% or less of genes are paralogs) blocks. This threshold was chosen because it classifies a given block as paralogous already when at least two or three paralogs are present. Indeed, the tendency of paralogous blocks to show higher gene density regime and a lower median intergenic distance than random blocks was observed across every block. This is particularly prominent in BLCA novel microsynteny (Fig. 2A, Supplementary Fig. 7).

Conversely, the density dynamics in non-paralogous microsyntenic blocks is more diverse. The normalized gene density and median intergenic distance of the MLCA and PLCA novelties retained in extant animals do not differ from the density of randomly sampled blocks (Fig. 2A, Supplementary Fig. 7A). Most BLCA non-paralogous microsyntenic blocks retained in ecdysozoans and lophotrochozoans, on the other hand, follow a high density regime. Not only is their density higher than that of random blocks (p -value Wilcoxon rank-sum test <0.05 , Fig. 2A, Supplementary Fig. 5), the intergenic distance of consecutive syntenic genes is also generally lower (Supplementary Fig. 7A, p -value Wilcoxon rank-sum test <0.05). In addition, most lophotrochozoan novel blocks also follow a high gene density regime (Fig. 2A, Supplementary Fig. 5). Interestingly, most non-paralogous blocks that emerged in the stem leading to the last common ancestor of jawed vertebrates displayed a low density (Fig. 2A, Supplementary Fig. 5), and high intergenic distance (Supplementary Fig. 7), suggesting most of these blocks follow a low gene density regime. This result is consistent with the observation that the blocks conserved within amniotes exhibit a lower gene density than the rest of the genome [16].

Our data can be used to provide thresholds for estimating whether any block follows a high or low density regime, without sampling random blocks. The lower and higher quartiles of normalized gene density of the random blocks of amphioxus (*B. lanceolatum*, 1.47 and 1.84, respectively) and scallop (*M. yessoensis*, 1.34 and 1.98, respectively), provide tentative thresholds for high and low density regimes in those two species. Accordingly, if any block has a normalized gene density below 1, it is likely to follow a low density regime, whereas any block with a normalized gene density higher than 2 is more likely to follow a high density regime. To further facilitate exploration of the density regimes of microsyntenic

blocks, we developed a new browser <http://synteny.csb.univie.ac.at/>.

In summary, paralogous blocks are in most cases more closely packed together than randomly sampled blocks, consistent with the fact that regions that underwent tandem duplications are associated with a higher gene density [46, 47]. Interestingly, the density of non-paralogous blocks relative to the rest of the genome differs between taxa (Fig. 2A). While most MLCA and PLCA novelties display average gene density, most of the BLCA novelties retained in lophotrochozoans and ecdysozoans exhibit a high density regime. A majority of lophotrochozoan novel microsyntenic blocks follow a high density regime as well, while the majority of jawed vertebrates microsyntenic blocks follow a low density regime.

Maintenance of density regimes in orthologous blocks

We asked whether the observed difference in normalized gene densities is a result of contraction or expansion of microsyntenic blocks or from the preferential retention of blocks with high or low density.

In order to compare density regimes of orthologous blocks across taxa, we calculated the relative “change of normalized gene density” (CNGD, [Methods](#)). This value indicates whether blocks are likely to follow a low (CNGD $<<0$), high (CNGD $>>0$) or average (CNGD ≈ 0) gene density regime. The comparison of all the possible taxon pairs revealed that the density regime of orthologous MLCA, PLCA and BLCA novel blocks were positively correlated in most cases (Fig. 2B, Supplementary Fig. 6), implying evolutionary conservation of gene density regimes.

As the density of most MLCA and PLCA novel blocks does not differ significantly from that of random blocks (Fig. 2A, Supplementary Fig. 5), yet correlates positively across taxa (Fig. 2B), we posit that these microsyntenic blocks could still experience selective pressures to maintain their density.

Since lophotrochozoans retained a particularly high number of the BLCA novel blocks (93%, Fig. 1B) they can be used as a reference to compare density regime conservation across all bilaterian clades. The majority of the lophotrochozoan BLCA novel blocks follow a high gene density regime (Fig. 2A, Supplementary Fig. 5). Moreover, lophotrochozoan BLCA blocks mostly follow the same density regimes as their orthologs in ambulacrarians (82.4% of BLCA blocks), vertebrates (14.8% of BLCA blocks) and ecdysozoans (44.1% of BLCA blocks) (Fig. 2B, C, Supplementary Fig. 8).

These results suggest that a substantial proportion of the microsyntenic blocks that emerged at the bilaterian stem followed a high density regime, and this ancestral regime was retained across extant taxa. In addition, most

MLCA and PLCA blocks follow average genome density, a property conserved across taxa. The maintenance of density regimes across bilaterians suggests that most MLCA, PLCA and BLCA blocks did not undergo contraction or expansion in any lineages.

Gene density predicts coexpression properties of microsyntenic genes

We next aimed to determine the impact of different microsyntenic density regimes onto gene expression. Given the proposed link between gene density and coexpression [14, 15], we hypothesized that denser microsyntenic blocks would exhibit a higher coexpression of their genes. To test this hypothesis, we used transcriptomic data (developmental series or tissues) from two lophotrochozoans (*Crassostrea gigas* and *Mizuhopecten yessoensis*), two ambulacrarians (*Strongylocentrotus purpuratus* and *Saccoglossus kowalevskii*) and two jawed vertebrates (*Mus musculus* and *Callorhinchus milii*). In line with a previous study [13], we define the coexpression levels of genes within a syntenic block as the mean Spearman correlation coefficient among all pairs of expressed genes of the block (which we will refer to as “block coexpression”, see Methods, Supplementary Fig. 2D).

We then compared the coexpression of paralogous and non-paralogous blocks to that of randomly sampled blocks. We found that higher density blocks exhibit a higher coexpression than random blocks (Fig. 3, Supplementary Fig. 9). This suggests that the constraints to maintain the coexpression of genes drives the maintenance of high density blocks. Conversely, the block coexpression of vertebrate microsyntenic novelties, enriched in low density blocks, did not differ significantly from

that of random blocks. As the expression of genes within vertebrate GRBs is not significantly correlated [12], this suggests that vertebrate microsyntenic novelties might be conserved more due to GRB [6, 9, 10, 12] rather than coexpression constraints.

We investigated gene density and coexpression properties of microsynteny to test their predictability of coregulation in several well studied microsynteny. We first examined the Wnt5–7 block, which is comprised of two conserved syntenic pairs (*fbxl14-wnt5* and *atxn10-wnt7*, identified by [6]), as well as other known neighbors of *wnt5* and/or *wnt7* (e.g. *erc1/2*, *cacna2d*, [48, 49]). The Wnt5–7 block showed a normalized gene density significantly lower than that of randomly sampled blocks in most of the bilaterians we investigated (Fig. 4A, B, Supplementary Fig. 10A). Parsimony suggests that its gene density was low in the BLCA, and did not undergo contraction in any lineage. The fact that the reciprocal of gene density of the Wnt5–7 block scales at the same rate as the genome-wide average density, and is almost always of lower density than average (Fig. 4D) further supports this scenario. The presence of multiple functionally unrelated genes points to the absence of constraints on the coexpression of the block components. This is consistent with the hypothesis that *wnt7* and *atxn10* form a GRB [6]: The bystander gene (*atxn10*) contains regulatory elements in its introns that target the transdev gene (*wnt7*), resulting in a cis-regulatory constraint driving the conservation of their microsyntenic association.

Even though the vast majority of blocks retained their ancestral density regime across the investigated metazoans, there is one notable exception: the Hox cluster. While the Hox cluster shows a lower density than random blocks

		Blocks composed of non-paralogous genes						Blocks composed of paralogous genes					
		Meta	Para	Planu	Bila	Vert	Loph	Meta	Para	Planu	Bila	Vert	Loph
Normalized gene density	<i>M. musculus</i>	0	0	0	1	31		0	0	1	4	0	
	<i>C. millii</i>	0	0	1	4	173		1	0	1	4	4	
	<i>S. kowalevskii</i>	5	1	52	129			10	2	29	53		
	<i>S. purpuratus</i>	6	1	27	55			21	2	45	50		
	<i>M. yessoensis</i>	4	0	22	102		56	8	2	27	20		9
	<i>C. gigas</i>	1	1	26	66		43	14	2	30	20		11
Block coexpression	<i>M. musculus</i>	0	0	0	1	28		0	0	1	2	0	
	<i>C. millii</i>	0	0	1	4	173		1	0	1	4	4	
	<i>S. kowalevskii</i>	5	1	52	128			9	0	26	39		
	<i>S. purpuratus</i>	6	1	27	55			21	2	45	50		
	<i>M. yessoensis</i>	4	0	22	102		56	7	2	27	20		8
	<i>C. gigas</i>	1	1	25	66		43	13	2	29	20		11

* n.s.

□ no block detected

■ n/a

■ Observed value lower than random

■ Observed value higher than random

Fig. 3 Higher coexpression in gene-dense microsyntenic blocks. Table noting the number of blocks inherited from key nodes (columns) that are found in bilaterian species (rows). The upper half of the table shows how the density of observed blocks differs from randomly sampled blocks, whereas the lower part shows how the block coexpression of observed microsynteny differs from randomly sampled blocks. Colors highlight how the observed blocks with a median density/coexpression differs from that of randomly sampled blocks (see legend). Vivid colors note a significant difference (p-value < 0.05, Wilcoxon rank-sum test)

in most bilaterians (particularly significant in lophotrochozoans and ecdysozoans, Fig. 4A, C, Supplementary Fig. 10B), its density is significantly higher than random in vertebrates. In addition, the reciprocal of gene density of most invertebrate Hox clusters scales linearly with assembly size, at a similar rate as the genome-wide average, even though the density of these clusters is generally lower than the genome-wide average (Fig. 4E). In vertebrates, however, the Hox cluster does not scale at the same rate as the genome-wide average (Fig. 4E), and most Hox clusters are at least twice as dense as this average. Consistent with previous observations [50, 51], our findings show that the Hox cluster was less dense than the rest of the genome in the BLCA, and contracted before the vertebrate LCA.

Apart from coexpression, gene density regimes might also reflect changes in genome topology. As noted above, anecdotal evidence suggest that high- and low-density regions can be respectively associated with the boundaries and the interior of TADs in *Drosophila* [44]. Consistently, the gene-dense HoxD cluster of vertebrates is located at the interface of two TADs, while the low density amphioxus Hox cluster is located within a single TAD [11]. However, in order to test whether high-density microsynteny are associated with specific topological properties, future efforts should be directed towards the generation of high resolution Hi-C data in a broad spectrum of invertebrate species. To this end, our density regimes of microsynteny may be useful as validators and predictors of the topological properties.

Conclusions

This study provides evidence that the evolutionary transition towards bilaterality in animals coincides with the emergence of a large set of microsynteny, the majority of which are still retained in many extant cnidarian and bilaterian genomes. Our results suggest that microsyntenic blocks exist in different density regimes across animal genomes. In particular, syntenies that originated in the BLCA are generally associated with higher gene densities and display a higher level of gene coexpression than the remainder of the genome. While their topological

organization remains to be investigated, this provides additional evidence for the proposed emergence of CTCF-driven topologically associating domain organization in bilaterians [42]. It should however be noted that a substantial amount of PLCA novel microsynteny have been maintained in cnidarians, which implies the existence of additional constraints underlying the maintenance of microsyntenic blocks. We also reveal that the constraints on gene density within blocks are heterogeneous depending on the node of emergence of the blocks and that gene density is correlated with microsyntenic gene coexpression. In contrast, coexpression does not seem to drive the maintenance of low density syntenies, hinting towards other types of gene regulation [6, 28]. Apart from the known distinct regulatory regimes in Hox clusters, little data exist to speculate about potential fundamental functional differences in how vertebrate and invertebrate genomes are regulated, constrained, and organized. Microsyntenic profiling together with further regulatory genomic information on topological organization and gene regulation across metazoans will thus help to better understand the regulatory causes or consequences of the ancient syntenic transitions during animal evolution.

Methods

List of software and scripts

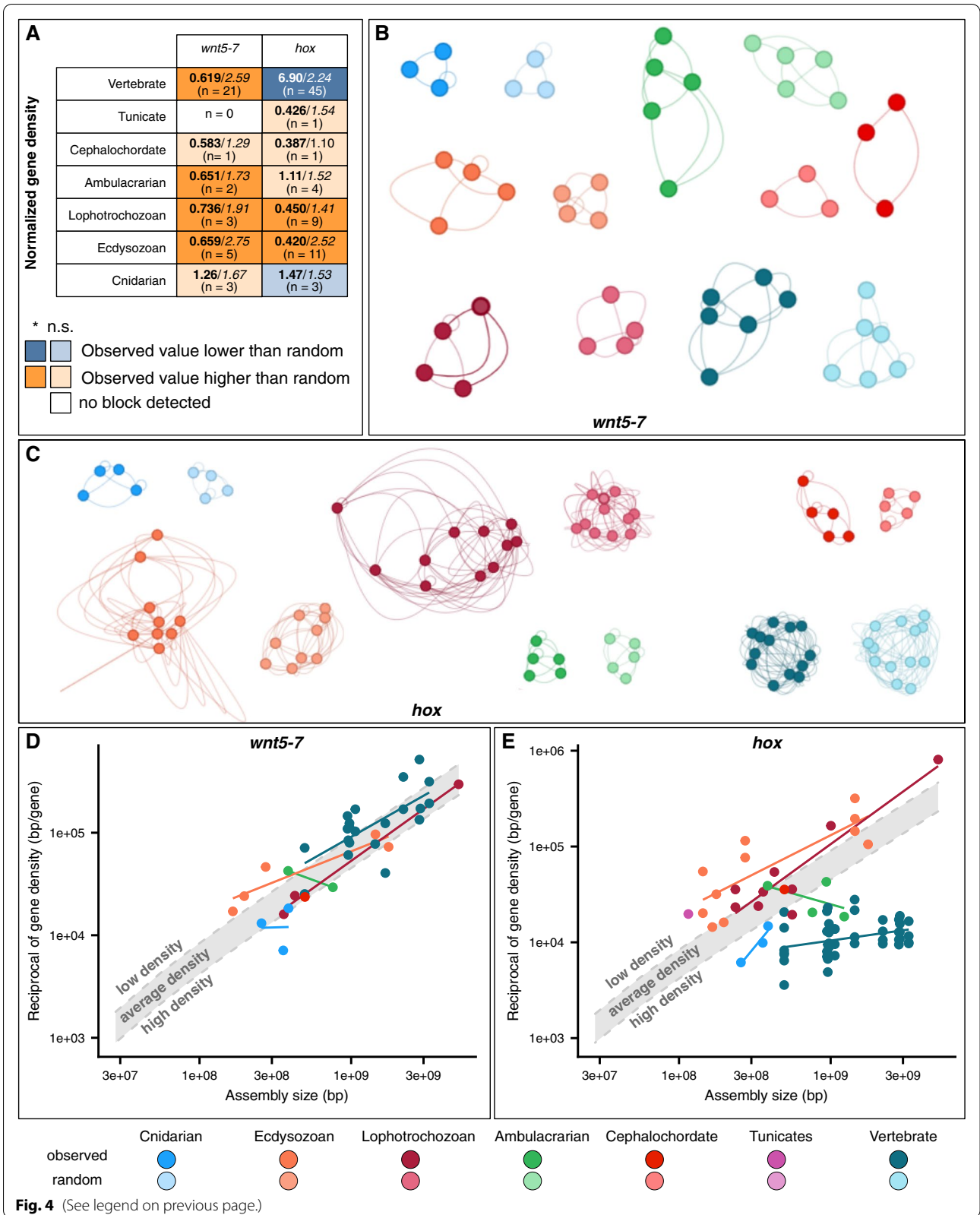
List of software used can be found in the Additional File 1. The repository containing all the details on commands, scripts, and datasets used in this study at the time of submission is provided in pdf format as Supplementary Dataset 1. An online version is also available at https://bitbucket.org/N_Robert/syntenic-density-and-transitions/.

Orthology assignment

Proteomes, annotations, and genomes of 38 species were downloaded from NCBI (*Homo sapiens*, *Mus musculus*, *Danio rerio*, *Gallus gallus*, *Xenopus tropicalis*, *Callorhynchus milii*, *Acropora millepora*, *Ciona intestinalis*, *Latimeria chalumnae*, *Maylandia zebra*, *Lepisosteus*

(See figure on next page.)

Fig. 4 Gene density dynamics of Hox and Wnt5–7 microsynteny across planulozoans. **A** Normalized gene density (observed/random, n is the number of blocks) in Wnt5–7 and Hox microsyntenic blocks (including duplicates and/or split ones) found in extant metazoan taxa (rows). Colors note whether median normalized gene density of the block is higher in observed or randomly sampled blocks (see legend). Vivid colors note a statistically significant difference ($p < 0.05$, Wilcoxon rank-sum test). Distributions of normalized gene density of both these blocks can be found in Supplementary Fig. 10. Wnt5–7 (**B**) and Hox (**C**) microsyntenic blocks displayed as graphs; nodes are gene families, edge lengths are the minimum normalized distance (distance in base pairs, normalized by assembly size) found between any given orthogroup pair within the taxonomic group. Self-edges are the minimum distance found between genes belonging to the same orthogroup. Reciprocal of raw gene density of Wnt5–7 (**D**) and Hox (**E**) as a function of assembly size, using a log-log scale. The upper and lower limits of the gray band correspond to the regression line explaining reciprocal gene density as a function of assembly size (Supplementary Fig. 3), and the regression line explaining reciprocal of the double of gene density as a function of assembly size, respectively. Both these values correspond to the tentative thresholds of 1 and 2 of the normalized gene density for identifying high and low gene density regimes, respectively



oculatus, *Chelonia mydas*, *Parasteatoda tepidariorum*, *Exaiptasia pallida*, *Hippocampus comes*, *Acanthaster planci*), Ensembl Metazoa Release 45 (*Adineta vaga*, *Amphimedon queenslandica*, *Anopheles gambiae*, *Branchiostoma lanceolatum*, *Caenorhabditis elegans*, *Capitella teleta*, *Crassostrea gigas*, *Daphnia pulex*, *Drosophila melanogaster*, *Helobdella robusta*, *Ixodes scapularis*, *Lingula anatina*, *Lottia gigantea*, *Nematostella vectensis*, *Strigamia maritima*, *Strongylocentrotus purpuratus*, *Tribolium castaneum*, *Trichoplax adhaerens*, *Salpingoeca rosetta*, *Capsaspora owczarzewski*), or the National Human Genome Research Institute (*Hydra vulgaris*, *Mnemiopsis leidyi*). The *Clytia hemisphaerica* data was retrieved from <http://marimba.obs-vlfr.fr>. The *Aurelia aurita* genome, annotation and proteome were provided by the authors of [52]. The genome of *Hofstenia miamia* and its annotation were retrieved from <http://srivastavalab.rc.fas.harvard.edu>. The genome, proteome, and annotation of *Saccoglossus kowalevskii* and *Ptychodera flava* were downloaded from OIST (<https://groups.oist.jp/molgnu/hemicordate-genomes>), the ones of *Schmidtea mediterranea* from planmine (<http://planmine.mpi-cbg.de>), and the ones of *Hoilungia hongkongensis* from <https://bitbucket.org/molpalmuc/hoilungia-genome/src/master/tracks/>. The CDS sequences and genomes of *Pleurobrachia bachei* and *Sycon ciliatum* were retrieved, respectively, from Neurobase (<https://neurobase.rc.ufl.edu>), and DataDryad (<https://datadryad.org/resource/doi:10.5061>). The mapping of the transcripts of these two species onto their genomes was done using gmap. The genome and annotation of *Mizuhopecten yessoensis* was provided by the authors of [53]. *Euprymna scolopes* chromosome-level assembly and annotation was provided by the authors of Belcaid et al. [7]. The proteins of the 49 aforementioned species were assigned to orthogroups using the Orthofinder v2.3 algorithm [54], based on the results of an all-against-all BLASTP with an e-value threshold of 0.001. In order to avoid erroneous ortholog inference due to ambiguous tree topologies, all the genes within the same orthogroup were considered to be orthologs for all subsequent analyses.

Inference of microsyntenic blocks and block emergence nodes

Syntenic blocks were inferred using in-house scripts, as described in [1, 2] (See Fig. 1A). These scripts can be found at https://github.com/nijibabulu/metazoan_syteny. All-against-all pairwise microsyntenic blocks were first built, requiring to comprise at least three orthologs, separated by no more than five intervening genes, but collinearity was not required. The minimum of three genes was chosen to increase sensitivity, as the use of scaffold-level assemblies do not allow the detection of large syntenic blocks. With

a maximum of five intervening genes, the distribution of the number of intervening genes between consecutive orthologs in the block follows a power law [13]. This also limits the number of false positives, as an increase of the allowed number of intervening genes results in an increase in the number of false positives [1]. The pairwise syntenic blocks were then fused into multi-species blocks, if they shared at least three genes, or if there are more than three genes in the pairwise block, at least 50% of the orthologs. The multi-species microsyntenic blocks were then filtered according to the node where we inferred them to be present (Supplementary Fig. 2B). For each node, we define an ingroup as all the species descending from this node, and the outgroup as all the species outside of the ingroup. The children ingroups are one level above the nodes (e.g., acoel, deuterostomes and protostomes are the children clades of the LCA of Bilateria). We define a novel microsyntenic block as one found in two or more species of two or more children ingroup but not found in the outgroup (Supplementary Fig. 2B). An inherited microsyntenic block is a block found in at least two species of the ingroup and at least two species of the outgroup. In order to evaluate the number of blocks that we would detect by chance, we ran our microsyteny pipeline on three independent permutations of the genomes. We considered blocks to be lost in a taxonomic group if no species of the group was present in the multi species block. We used this method on the actual genomes and three iterations of shuffled genomes, where the relative position of genes on the genome was shuffled.

Gene density analysis

The gene density within a given block was calculated as the number of genes per base pairs of a block. The boundaries of a block are defined as the outermost coding bases of genes exhibiting conserved syteny. This measure was used for the sake of consistency across all included species, as the coordinates of untranslated regions of the mRNAs were absent of some of the genome annotations we used. The number of genes includes all the genes of which both start and stop codons are located within the block boundaries. The gene density per block was then normalized by the gene density within the whole genome (total number of genes divided by total number of base pairs of the assembly). We refer to this value as normalized gene density. Since the reciprocal of whole genome density increases linearly with the assembly size (Supplementary Fig. 4) this normalization allows us to group normalized gene density measures from animals with different genome sizes according to their clade (see Supplementary Fig. 1).

We also calculated the median intergenic distance of consecutive syntenic orthologs. For each block, every distance (in base pairs) between consecutive orthologs

was calculated (distance between boundaries of coding regions of the two genes), and the median value of each block was retained.

Each block was defined as non paralogous or paralogous. We define a paralogous block as comprising more than 40% of genes belonging to the same orthogroup (the region underwent several events of segmental duplications). Non paralogous blocks are all the other blocks that do not satisfy this criteria.

For each observed block, 100 blocks were randomly sampled. Randomly sampled blocks were built using the methods described in [13]. A detailed description of the methods can also be found in Supplementary Fig. 2C. Differences in the normalized gene density between the observed and randomly sampled blocks were assessed using a Wilcoxon rank-sum test and a threshold of significance of 0.05. For each multi species block found in at least two taxa, we also calculated the change of normalized gene density relative to random (CNGD), such as $CNGD = (\text{median}(\text{normalized_gene_density}_{\text{obs}}) - \text{median}(\text{normalized_gene_density}_{\text{rand}})) / \text{median}(\text{normalized_gene_density}_{\text{rand}})$. Accordingly, if the median normalized gene density of a multi species block ($\text{median}(\text{normalized_gene_density}_{\text{obs}})$) is twice as dense as that of their randomly sampled counterparts ($\text{median}(\text{normalized_gene_density}_{\text{rand}})$), this will result in $CNGD = 1$. If the observed blocks are of a density that is half of the random blocks, this will result in $CNGD = -1$. If the median normalized gene density of observed and randomly sampled blocks is identical, then $CNGD = 0$.

Annotation of the genes of the Hox and Wnt5–7 clusters

If available, annotations of *hox* genes were recovered from the literature, and their orthologs in our taxonomic sample were identified by reciprocal BLAST searches and phylogenies. The *hox* queries were from *Latimeria chalumnae* [55], *Homo sapiens*, *Mus musculus*, *Danio rerio*, *Callorhinchus milii*, *Chelonia mydas*, *Ciona intestinalis*, *Strongylocentrotus purpuratus* [27, 56], *Euprymna scolopes* [7], *Hofstenia miamia* [57], *Schmidtea mediterranea* [58], *Nematostella vectensis* [59], *Caenorhabditis elegans* [60], *Clytia hemisphaerica* [61], *Parasteatoda tepidariorum* [62], *Anopheles gambiae*, *Drosophila melanogaster*, *Tribolium castaneum*, *Strigamia maritima*, *Ixodes scapularis* [63], *Capitella teleta*, *Lottia gigantea*, *Daphnia pulex* [1], *Saccoglossus kowalevskii*, *Ptychodera flava* [2], *Mizuhopecten yessoensis* [53]. Among the microsyntenic blocks identified by our pipeline, one block (multi-species block id 953, see Supplementary Dataset 3) comprised the following genes: *wnt5*, *wnt7*, *fbxl14*, *atxn10*, *erc1/2*, *cacna1d*, *cacna2d*, *ninj1/ninj2* and *dcp1a/dcp1b*. This block is referred to as the Wnt5–7 block throughout the manuscript. We identified the orthologs of the genes

in all the species of our dataset by reciprocal BLAST searches and phylogenies. Only the Hox and Wnt5–7 clusters comprising at least three genes, separated by no more than five intervening genes were retained.

Functional annotation and GO term enrichment analysis

Annotation and GO annotation of the proteomes of all species was done using eggNOG-mapper version 2 [64]. GO enrichment analysis was done using the GOATOOLS API [65]. The GO enrichment was done without propagating GO counts to parent GOs, using a threshold of 0.05 for Benjamini-Hochberg corrected *p*-values. GO terms enriched in the MLCA novel blocks were required to be enriched in at least eight metazoan species, distributed in at least two metazoan ingroups (i.e., Porifera, Ctenophora, and Parahoxozoa). GO terms enriched in the PLCA novel blocks were required to be enriched in at least three cnidarians and eight bilaterian species. GO terms enriched in the BLCA novel blocks were required to be enriched in at least four protostomes and four deuterostomes species. We provide the tables listing the GO terms enriched in MLCA (Supplementary Table 1), PLCA (Supplementary Table 2) and BLCA (Supplementary Table 3) blocks.

Block coexpression analysis

The methods used for determining transcripts abundance in *Strongylocentrotus purpuratus*, *Saccoglossus kowalevskii*, *Mizuhopecten yessoensis* and *Crassostrea gigas* were described in Zieger et al. [66]. In line with this study, the transcript abundances in *Mus musculus* [67] and *Callorhinchus milii* [68] tissues were quantified as TPMs (Transcript per Kilobase per Million of reads) using Kallisto [69] with default settings for paired end reads. For each block, we calculated the block co expression value as described in Zimmerman et al. [13]. For each block comprising at least three expressed genes, we calculated the Spearman correlation of expression of every combination of expressed genes. To allow the averaging of correlations [70], the array of correlations of a given block was then transformed using the inverse hyperbolic tangent function (Fisher transformation), the mean of the Fisher-transformed correlation was then calculated, and this value was then transformed into a correlation using the hyperbolic tangent function (reverse Fisher transformation). In addition, if an untransformed correlation value was -1 or 1 , we transformed it into the next possible floating point value towards zero using the `numpy.nextafter()` function. This was done as the inverse hyperbolic tangent of 1 is infinite, and that of -1 is negative infinite.

Synten browser

We introduce a novel browser containing our complete data set which allows users to investigate the syntenic density relationships of orthogroups with their genes of interest. In order to present a better visual comparison, we normalized the minimum base pair distances between each syntenic orthogroup pair of each multi-species block that we identified, by the total genome assembly size of each species. Our browser has four panels: while the first panel displays the boxplots of distance distributions for each taxonomic node, the second one displays the geometric densities of these; the third panel illustrates the orthogroups and the base pair distances between them in a network graph, and the fourth panel is a dictionary of all genes and their functions. Each panel is user-interactive and it is also possible to compare our observed values with the randomized cases. Users can first search for genes or annotations of their interest that exist in our dataset to find the multi-species block ID and the taxonomic group information, and then navigate to that block. Our browser is available at: <http://synten.csb.univie.ac.at/>.

Abbreviations

GRB: Gene Regulatory Block; BLCA: Bilaterian Last Common Ancestor; PLCA: Planulozoan Last Common Ancestor; MLCA: Metazoan Last Common Ancestor; GO: Gene Ontology; TAD: Topologically Associating Domain; CTCF: CCCTC-binding factor; CNGD: Relative change of normalized gene density; TPM: Transcript per Kilobase per Million of reads.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08304-2>.

Additional file 1: Supplementary Figure 1. List of software used and Supplementary Figures 1 to 10.

Additional file 2: Supplementary Dataset 1. Code and scripts from the repository.

Additional file 3: Supplementary Dataset 2. GO enrichment tables.

Additional file 4: Supplementary Dataset 3. Syntenic density database.

Additional file 5: Supplementary Table 1. Summary of GO terms enriched in MLCA blocks

Additional file 6: Supplementary Table 2. Summary of GO terms enriched in PLCA blocks.

Additional file 7: Supplementary Table 3. Summary of GO terms enriched in BLCA blocks.

Acknowledgments

The authors thank Elisabeth Zieger and Elizabeth Hambleton for critically reading an early version of this manuscript.

Authors' contributions

N.S.M.R., A.M., C.R.V. and O.S. designed research; N.S.M.R., F.S., B.Z., O.S. performed research; N.S.M.R., A.M., C.R.V. and O.S. analyzed data and wrote the paper. All authors read and approved the final manuscript.

Funding

N.S.M.R., F.S., and O.S. were supported by the Austrian Science Fund grant P32190. The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The repository containing all the details on commands, scripts, and datasets used in this study at the time of submission is provided in pdf format as Supplementary Dataset 1. An online version is also available at https://bitbucket.org/N_Robert/syntenic-density-and-transitions/.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no conflict of interest.

Author details

¹Department of Neurosciences and Developmental Biology, University of Vienna, Althanstrasse 14, 1090 Wien, Austria. ²Department of Biology, University of Konstanz, 78457 Constance, Germany.

Received: 21 July 2021 Accepted: 10 January 2022

Published online: 17 February 2022

References

1. Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U, et al. Insights into bilaterian evolution from three spiralian genomes. *Nature*. 2013;493(7433):526–31.
2. Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, et al. Hemichordate genomes and deuterostome origins. *Nature*. 2015;527(7579):459–65.
3. Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, et al. The Trichoplax genome and the nature of placozoans. *Nature*. 2008;454(7207):955–60.
4. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, et al. Sea Anemone genome reveals ancestral Eumetazoan gene repertoire and genomic organization. *Science*. 2007;317(5834):86–94.
5. Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature*. 2008;453(7198):1064–71.
6. Irimia M, Tena JJ, Alexis MS, Fernandez-Minan A, Maeso I, Bogdanovic O, et al. Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res*. 2012;22(12):2356–67.
7. Belcaid M, Casaburi G, McAnulty SJ, Schmidbauer H, Suria AM, Moriano-Gutierrez S, et al. Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. *Proc Natl Acad Sci U S A*. 2019;116(8):3030–5.
8. Zhao T, Zwaenepoel A, Xue J-Y, Kao S-M, Li Z, Schranz ME, et al. Whole-genome microsynteny-based phylogeny of angiosperms. *Nat Commun*. 2021;12(1):3498.
9. Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engstrom PG, Fredman D, et al. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res*. 2007;17(5):545–55.
10. Engstrom PG, Ho Sui SJ, Drivenes O, Becker TS, Lenhard B. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res*. 2007;17(12):1898–908.
11. Acemel RD, Tena JJ, Irastorza-Azcarate I, Marlétaz F, Gómez-Marín C, de la Calle-Mustienes E, et al. A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. *Nat Genet*. 2016;48(3):336–41.
12. Akalin A, Fredman D, Arner E, Dong X, Bryne J, Suzuki H, et al. Transcriptional features of genomic regulatory blocks. *Genome Biol*. 2009;10(4):R38.
13. Zimmermann B, Robert NSM, Technau U, Simakov O. Ancient animal genome architecture reflects cell type identities. *Nat Ecol Evol*. 2019;3(9):1289–93.

14. Sun M, Zhang J. Chromosome-wide co-fluctuation of stochastic gene expression in mammalian cells. *PLoS Genet.* 2019;15(9):e1008389.
15. Ng YK, Wu W, Zhang L. Positive correlation between gene coexpression and positional clustering in the zebrafish genome. *BMC Genomics.* 2009;10(1):42.
16. Larkin DM, Pape G, Donthu R, Auviel L, Welge M, Lewin HA. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res.* 2009;19(5):770–7.
17. Feuillet C, Keller B. High gene density is conserved at syntenic loci of small and large grass genomes. *Proc Natl Acad Sci U S A.* 1999;96(14):8265–70.
18. Lv J, Havlak P, Putnam NH. Constraints on genes shape long-term conservation of macro-synteny in metazoan genomes. *BMC Bioinformatics.* 2011;12(Suppl 9):S11.
19. Dunn CW, Giribet G, Edgecombe GD, Hejnol A. Animal phylogeny and its evolutionary implications. *Annu Rev Ecol Evol Syst.* 2014;45(1):371–95.
20. Telford MJ, Moroz LL, Halanych KM. A sisterly dispute. *Nature.* 2016;529(7586):286–7.
21. Philippe H, Poustka AJ, Chiodin M, Hoff KJ, Dessimoz C, Tomiczek B, et al. Mitigating anticipated effects of systematic errors supports sister-group relationship between Xenacoelomorpha and Ambulacraria. *Curr Biol.* 2019;29(11):1818–1826.e6.
22. Cannon JT, Vellutini BC, Smith J, Ronquist F, Jondelius U, Hejnol A. Xenacoelomorpha is the sister group to Nephrozoa. *Nature.* 2016;530(7588):89–93.
23. Dos Reis M, Thawornwattana Y, Angelis K, Telford MJ, Donoghue PCJ, Yang Z. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr Biol.* 2015;25(22):2939–50.
24. Whelan NV, Kocot KM, Moroz LL, Halanych KM. Ctenophore relationships and their placement as the sister group to all other animals. *Nat Ecol Evol.* 2017;1:1737–46.
25. Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Franco AD, et al. A large and consistent Phylogenomic dataset supports sponges as the sister group to all other animals. *Curr Biol.* 2017;27(7):958–67.
26. Sacerdot C, Louis A, Bon C, Berthelot C, Roest CH. Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol.* 2018;19(1):166.
27. Kuraku S, Meyer A. Detection and phylogenetic assessment of conserved Synteny derived from whole genome duplications. In: Anisimova M, editor. *Evolutionary genomics* [internet]. Totowa, NJ: Humana Press; 2012. p. 385–95. Available from: http://link.springer.com/10.1007/978-1-61779-582-4_14.
28. Touceda-Suárez M, Kita EM, Acemel RD, Firbas PN, Magri MS, Naranjo S, et al. Ancient genomic regulatory blocks are a source for regulatory gene deserts in vertebrates after whole-genome duplications. *Nowick K, editor. Mol Biol Evol.* 2020;37(10):2857–64.
29. Cañestro C, Catchen JM, Rodríguez-Marí A, Yokoi H, Postlethwait JH. Consequences of lineage-specific gene loss on retinoic acid signaling in vertebrate genomes. *Gojobori T, editor. PLoS Genet.* 2009;5(5):e1000496.
30. Blackmon H, Ross L, Bachtrog D. Sex determination, sex chromosomes, and karyotype evolution in insects. *JHERED.* 2017;108(1):78–93.
31. Raible F. Vertebrate-Type Intron-Rich Genes in the Marine Annelid *Platynereis dumerilii*. *Science.* 2005;310(5752):1325–6.
32. Fernández R, Gabaldón T. Gene gain and loss across the metazoan tree of life. *Nat Ecol Evol.* 2020;4(4):524–33.
33. Guijarro-Clarke C, Holland PWH, Paps J. Widespread patterns of gene loss in the evolution of the animal kingdom. *Nat Ecol Evol.* 2020;4(4):519–23.
34. Genikhovich G, Technau U. On the evolution of bilaterality. *Development.* 2017;144(19):3392–404.
35. Sebé-Pedrós A, Chomsky E, Pang K, Lara-Astiaso D, Gaiti F, Mukamel Z, et al. Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat Ecol Evol.* 2018;2(7):1176–88.
36. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, et al. The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature.* 2010;466(7307):720–6.
37. Grau-Bové X, Torruella G, Donachie S, Suga H, Leonard G, Richards TA. Dynamics of genomic innovation in the unicellular ancestry of animals. *eLife.* 2017;6:e26036.
38. Paps J, Holland PWH. Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. *Nat Commun.* 2018;9(1):1730.
39. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012;485(7398):376–80.
40. Gómez-Marín C, Tena JJ, Acemel RD, López-Mayorga M, Naranjo S, de la Calle-Mustienes E, et al. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc Natl Acad Sci U S A.* 2015;112(24):7542–7.
41. Sexton T, Cavalli G. The role of chromosome domains in shaping the functional genome. *Cell.* 2015;160(6):1049–59.
42. Heger P, Marin B, Bartkuhn M, Schierenberg E, Wiehe T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc Natl Acad Sci U S A.* 2012;109(43):17507–12.
43. Harmston N, Ing-Simmons E, Tan G, Perry M, Merckenschlager M, Lenhard B. Topologically associating domains are ancient features that coincide with metazoan clusters of extreme noncoding conservation. *Nat Commun.* 2017;8(1):441.
44. Hou C, Li L, Qin ZS, Corces VG. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol Cell.* 2012;48(3):471–84.
45. Lynch M, Conery JS. The origins of genome complexity. *Science.* 2003;302(5649):1401–4.
46. Zhang L, Lu HHS, Chung W, Yang J, Li W-H. Patterns of segmental duplication in the human genome. *Mol Biol Evol.* 2004;22(1):135–41.
47. Kono TJY, Brohammer AB, McGaugh SE, Hirsch CN. Tandem duplicate genes in maize are abundant and date to two distinct periods of time. *G3.* 2018;8(9):3049–58.
48. Kapasa M, Arhondakis S, Kossida S. Phylogenetic and regulatory region analysis of Wnt5 genes reveals conservation of a regulatory module with putative implication in pancreas development. *Biol Direct.* 2010;5(1):49.
49. Garriock RJ, Warkman AS, Meadows SM, D'Agostino S, Krieg PA. Census of vertebrate Wnt genes: isolation and developmental expression of *Xenopus* Wnt2, Wnt3, Wnt9a, Wnt9b, Wnt10a, and Wnt16. *Dev Dyn.* 2007;236(5):1249–58.
50. Duboule D. The rise and fall of Hox gene clusters. *Development.* 2007;134(14):2549–60.
51. Hoegg S, Boore JL, Kuehl JV, Meyer A. Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*. *BMC Genomics.* 2007;8(1):317.
52. Gold DA, Katsuki T, Li Y, Yan X, Regulski M, Ibberson D, et al. The genome of the jellyfish *Aurelia* and the evolution of animal complexity. *Nat Ecol Evol.* 2019;3(1):96–104.
53. Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol.* 2017;1(5):0120.
54. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238.
55. Amemiya CT, Alföldi J, Lee AP, Fan S, Philippe H, MacCallum I, et al. The African coelacanth genome provides insights into tetrapod evolution. *Nature.* 2013;496(7445):311–6.
56. Pascual-Anaya J, D'Aniello S, Kuratani S, Garcia-Fernández J. Evolution of Hox gene clusters in deuterostomes. *BMC Dev Biol.* 2013;13(1):26.
57. Brauchle M, Bilican A, Eyer C, Bailly X, Martínez P, Ladurner P, et al. Xenacoelomorpha survey reveals that all 11 animal Homeobox gene classes were present in the first Bilaterians. *Genome Biol Evol.* 2018;10(9):2205–17.
58. Currie KW, Brown DDR, Zhu S, Xu C, Voisin V, Bader GD, et al. HOX gene complement and expression in the planarian *Schmidtea mediterranea*. *Evodevo.* 2016;7(1):7.
59. DuBuc TQ, Ryan JF, Shinzato C, Satoh N, Martindale MQ. Coral comparative genomics reveal expanded Hox cluster in the cnidarian-Bilaterian ancestor. *Integr Comp Biol.* 2012;52(6):835–41.
60. Hench J, Henriksson J, Abou-Zied AM, Lüpbert M, Dethlefsen J, Mukherjee K, et al. The Homeobox genes of *Caenorhabditis elegans* and insights into their Spatio-temporal expression dynamics during embryogenesis. *PLOS ONE.* 2015;10(5):e0126947.
61. Leclère L, Horin C, Chevalier S, Lapébie P, Dru P, Peron S, et al. The genome of the jellyfish *Clytia hemisphaerica* and the evolution of the cnidarian life-cycle. *Nat Ecol Evol.* 2019;3(5):801–10.
62. Leite DJ, Baudouin-Gonzalez L, Iwasaki-Yokozawa S, Lozano-Fernandez J, Turetzek N, Akiyama-Oda Y, et al. Homeobox gene duplication and divergence in arachnids. *Mol Biol Evol.* 2018;35(9):2240–53.
63. Pace RM, Grbić M, Nagy LM. Composition and genomic organization of arthropod Hox clusters. *Evodevo.* 2016;7(1):11.
64. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through Orthology assignment by eggNOG-mapper. *Mol Biol Evol.* 2017;34(8):2115–22.
65. Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Warwick Vesztrocy A, Naldi A, et al. GOATOOLS: a Python library for gene ontology analyses. *Sci Rep.* 2018;8(1):10872.
66. Zieger E, Robert NSM, Calcino A, Wanninger A. Ancestral role of Ecdysis-related neuropeptides in animal life cycle transitions. *Curr Biol.* 2020;S0960982220315086.

67. Li B, Qing T, Zhu J, Wen Z, Yu Y, Fukumura R, et al. A comprehensive mouse transcriptomic BodyMap across 17 tissues by RNA-seq. *Sci Rep.* 2017;7(1):4200.
68. Peat JR, Ortega-Recalde O, Kardailsky O, Hore TA. The elephant shark methylome reveals conservation of epigenetic regulation across jawed vertebrates. *F1000Research.* 2017;6:526.
69. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(5):525–7.
70. Alexander RA. A note on averaging correlations. *Bull Psychon Soc.* 1990;28(4):335–6.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

