Genome analysis
SYNPHONI: scale-free and phylogeny-aware reconstruction of synteny conservation and transformation across animal genomes
Nicolas Serge Matthieu Robert 1,*, Fatih Sarigol1, Elisabeth Zieger 2 and Oleg Simakov 1,*
1Department of Neurosciences and Developmental Biology, University of Vienna, Vienna A-1030, Austria and 2Department of Evolutionary Biology, University of Vienna, Vienna A-1030, Austria
*To whom correspondence should be addressed.
Associate Editor: Can Alkan
Received on July 22, 2022; revised on September 24, 2022; editorial decision on October 18, 2022; accepted on October 19, 2022
Abstract
Summary: Current approaches detect conserved genomic order either at chromosomal (macrosynteny) or at subchromosomal scales (microsynteny). The latter generally requires collinearity and hard thresholds on syntenic region size, thus excluding a major proportion of syntenies with recent expansions or minor rearrangements. ‘SYNPHONI’ bridges the gap between micro- and macrosynteny detection, providing detailed information on both synteny conservation and transformation throughout the evolutionary history of animal genomes.
Availability and implementation: Source code is freely available at https://github.com/nsmro/SYNPHONI, implemented in Python 3.9.
Contact: nrobert.nicolas@orange.fr or oleg.simakov@univie.ac.at
Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction
Genes located on the same chromosome across species are considered syntenic [‘on the same ribbon’, (Renwick, 1971)]. A distinction is generally made between microsynteny (i.e. conservation of local/subchromosomal gene order) and macrosynteny (i.e. conserved synteny at chromosomal level). Both types of synteny are widely conserved across animal genomes (Irimia et al., 2012; Simakov et al., 2022), yet their functional significance remains elusive. While some microsyntenies are probably maintained because of constraints on the proximity between genes and their regulatory regions (Irimia et al., 2012; Kikuta et al., 2007; Zimmermann et al., 2019), little is known about the potential evolutionary advantages of macrosynteny (Sun and Zhang, 2019; Véron et al., 2011). Moreover, the distinction between micro- and macrosynteny can blur or even disappear in the context of 3D genomic structure and evolutionary modification. We thus developed SYNPHONI (detection of ancestral SYNTeny based on PHYlogeny and Ortholog Network Inference), a tool for detecting not only conserved synteny but also evolutionary changes to ancestral syntenic arrangements across all scales of metazoan genome organization and phylogenetic distance.

Many tools detect conserved microsyntenic blocks across multiple genomes [reviewed in (Lallemand et al., 2020)], but most require stringent conservation of gene order (collinearity) [e.g. MCSCANX (Wang et al., 2012), CYNTENATOR (Rödelsperger and Dieterich, 2010), AGORA (Nguyen, 2022), DESCHRAMLER (Kim et al., 2017) and i-ADHore (Proost et al., 2012)]. Exceptions include the Microsynteny tool (Simakov et al., 2013), Gecko3 (Winter et al., 2016) and EvolClust (Marcet-Houben and Gabaldon, 2019). However, these methods still require a priori specification of hard global thresholds on the number of intervening genes and syntenic block overlap. This imposes an artificial restriction on the scale at which synteny is considered informative, because it impedes the detection of blocks that are shared by smaller species subsets, blocks that underwent minor evolutionary modifications (e.g. expansions and translocations) or blocks that exceed a certain size [e.g. block shared by recently diverged taxa (Kaufmann and Frishman, 2014)].

In order to address the caveats outlined above, we designed SYNPHONI to work in a phylogeny-aware and scale-free manner. As such, SYNPHONI allows synteny reconstruction across different evolutionary distances, by outputting the syntenic complements of any given node of interest within the animal tree of life irrespective of collinearity or micro/macrosyntenic scale.

2 SYNPHONI algorithm
In SYNPHONI, a syntenic block is defined as a set of genes that are located in close proximity to each other (Supplementary Material, Step 2.5) in a given last common ancestor (LCA), without requiring...
SYNPHONI: reconstruct synteny evolution

5435

collinearity. The descendants of each ancestral block are then recovered in extant genomes, regardless of whether their proximity has been maintained or not.

SYNPHONI was implemented in Python 3.9. The command-line interface software accepts as input a file containing groups of orthologs (OGs) files containing gene coordinates and a species cladogram with phylogenetic clade names as node labels (polymorphisms allowed).

The SYNPHONI algorithm proceeds as follows: build a matrix comprising the smallest number of intervening genes (d) between every possible pair of OGs (including self-pairs, i.e. syntenic paralogs) for all the species of the dataset. Only OG pairs that are syntenic in two or more species are kept. For each phylogenetic node of interest (N), we define its ingroup clades (e.g. Fig. 1A species 2/3 and 4/5 are descendants of N), its sister group (e.g. Fig. 1A species 5 is the sister group to the ingroup clades of N) and its outgroup (all species of the tree excluding the ingroup clades). The synteny of an OG pair (OGi, OGj) is stated in N, if it is detected as syntenic in a required minimum number of phylogenetic clades (see Supplementary Material for details).

SYNPHONI returns the number of phylogenetic clades (k) where an OG pair is syntenic. For each clade (g), the ancestral distance dN(OGi-OGj) is then calculated as the mode of the kernel density estimate of the distances between all syntenic OGi and OGj pairs (Supplementary Material). The ancestral distance of OGi and OGj in the node of interest N (noted dN(OGi, OGj)) is thus defined as:

\[ d_N(OG_i, OG_j) = \frac{1}{k} \sum_{k=1}^{d} d_N(OG_i, OG_j) \]

Next, an OG network is built (raw graph), with edges stating synteny in N and with dN as an edge attribute. Only edges with dN < nmax are kept and the connected components are isolated. Details on the optimization of the nmax parameter can be found in the Supplementary Material. In brief, nmax ensures that SYNPHONI respects the original definition of a syntenic block (Pevzner and Tesler, 2003) by allowing a more or less limited degree of micro/macrosyntenic rearrangements depending on the evolutionary time frame (depth of N). Each connected component of the trimmed graph is further decomposed to isolate maximal cliques in the raw graph. This yields refined OGs corresponding to inferred ancestral macrosynteties present in N. Their descendants are recovered from extant genomes and the blocks sharing OGs are grouped into multi-species blocks. For final validation, the taxonomic composition of each multispecies block is assessed and only the blocks that were inherited from N are retained (Supplementary Material).

3 Insights and benchmarks

Running SYNPHONI on our dataset of 80 metazoan genomes (Supplementary Material), yielded 164 multispecies blocks in the metazoan LCA (Supplementary Fig. S4), 400 in the parazoan LCA, 616 in the planulozoan LCA, 893 in the nephrozoan LCA, 807 in the protostome LCA and 1153 in the LCA of coelate cephalopods. The analysis took <48 h using a single core.

For benchmarking, we compared SYNPHONI to the MicroSynteny tool (Smakov et al., 2013; Robert et al., 2022) and EvoClust (Marcot-Houben and Gabaldón, 2019), two methods that also detect non-collinear multispecies blocks across numerous genomes (Supplementary Material, Section S3). SYNPHONI provides the highest detection accuracy, as indicated by a significantly higher proportion of ‘core OGs’ (ancestral macrosynteny OGs) per multispecies block (Wilcoxon rank-sum tests, P < 0.05, Supplementary Fig. S7B). Contrary to other methods, SYNPHONI outputs not only a single set of blocks for all input species but identifies the multispecies blocks of all nodes of interest. Furthermore, it does not require hard thresholds on block size or the number of intervening genes but detects synteny regardless of micro/macroscale. This results in a much greater detection sensitivity (as indicated by a significantly higher number of species per multispecies block, Wilcoxon rank-sum tests, P < 0.01, Supplementary Fig. S7C). More importantly, however, it allows detailed insights into how synteny evolved across different time scales. As such, SYNPHONI can be used to reconstruct both the emergence and all major lineage-specific transitions of syntetic blocks across Metazoa (e.g. Fig. 1D–G, Supplementary Figs S9 and S10).

Taken together, SYNPHONI is an efficient, accurate, phylogenetic and, most importantly, scale-free alternative to

Fig. 1. SYNPHONI-based detection of evolutionary transitions in genetic linkages across Metazoa. (A–C) Graphical summary of the SYNPHONI pipeline. Colored circles correspond to different OGs, edge lengths in the networks are proportional to intergenic distance. (A) A network of inter-OG minimum distances is built for every species of the taxon sample, in order to infer the state of the ancestral OG network in any node of interest. (B) Ancestral macrosynteny units are isolated from the network, by deleting edges above a distance threshold (see Supplementary Material). (C) Ancestral microsynteny blocks are recovered from extant genomes, requiring only maintenance of synteny but not gene proximity (microsynteny). Each ancestral node corresponds to a chromosome, colored circles in gray genes, their color corresponds to their assigned OG. Distance between the circles corresponds to intergenic distance. (D–G) Colored squares represent members of the PIWIL2 (D, E) and the WNT5/7 (F, G) syntetic blocks detected in the nephrozoan LCA, as indicated in the respective legends. Genes are shown as microsyntetic (p35y), contiguous squares, if they are separated by no more than 10 intervening genes. For macrosyntetic genes (M5y, squares connected by a line), the number of intervening genes is provided in gray hexagons. Genes that are not located on the same chromosome are shown as non-syntenic (noSy, spaced squares). (D) SYNPHONI-based characterization of the PIWIL2 syntetic block for selected species (full dataset available in Supplementary Fig. S8), revealing frequent reshuffling of local gene order as well as species-specific losses of macro- or microsynteny. (E) Evolutionary scenario summarizing major syntetic transitions of the PIWIL2 cluster, including acquisition of MAT1/2 microsynteny in the last common ancestor of bilateria (+ΔSy) as well as microsynteny loss in echinoderms (−ΔSy) and macrosynteny loss in cephalopods and ecdysozoans (−M5y). (F) WNT5/7 gene cluster of selected species as identified by SYNPHONI (full dataset available in Supplementary Figs S8 and S9). As for the PIWIL2 cluster, reshuffling of local gene order and microsynteny losses are species-specific, whereas macrosynteny loss is gene-specific, indicating relaxed linkage constraints on FBLN1/2 and ATXN10. (G) Evolutionary scenario summarizing major syntetic transitions of the WNT5/7 cluster, such as acquisition of FBLN1/2 and ATXN10 microsynteny by the last common ancestor of bilateria (+ΔSy) as well as macrosynteny loss of FBLN1/2 in cephalopods and ecdysozoans (−M5y). Notably, these data indicate stronger linkage constraints on the evolutionary older macro- and microsyntetic relationships between WNT5/7, ERC1/2 and FBXL14, compared to the evolutionary younger relationships of ATXN10 and FBLN1/2. These examples demonstrate that SYNPHONI is a powerful tool for reconstructing the evolutionary history of genetic linkages at both micro- and macrosyntetic levels and across vast phylogenetic distances (A color version of this figure appears in the online version of this article).
other methods, inferring microsyntenic complements of deep nodes within the animal tree of life.

Acknowledgements
Computation was done at the Life Science Compute Cluster at the University of Vienna.

Funding
This work was supported by the Austrian Science Fund [P32190 to N.S.M.R., F.S. and O.S.].

Conflict of Interest: none declared.

Data availability
The data underlying this article are available in the article and in its Supplementary Material.

References
Irimia,M. et al. (2012) Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. Genome Res., 22, 2356–2367.
Kaufmann,S. and Frishman,D. (2014) Analysis of micro-rearrangements in 25 eukaryotic species pairs by SyntenyMapper. PLoS One, 9, e112341.
Kikuta,H. et al. (2007) Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res., 17, 545–555.
Kim,J. et al. (2017) Reconstruction and evolutionary history of eutherian chromosomes. Proc. Natl. Acad. Sci. USA, 114, E5379–E5388.
Lallemand,T. et al. (2020) An overview of duplicated gene detection methods: why the duplication mechanism has to be accounted for in their choice. Genes, 11, 1046.
Marcet-Houben,M. and Gabaldón,T. (2019) EvoClust: automated inference of evolutionary conserved gene clusters in eukaryotes. Bioinformatics, 36(4), 1265–1266. https://doi.org/10.1093/bioinformatics/btz706.
Nguyen,N. et al. (2022) Genomics in 2022: comparative tools for thousands of genomes and reconstructed ancestors. Nucleic Acids Research, 50, D1025–D1031. https://doi.org/10.1093/nar/gkab1191.
Pevzner,P. and Tesler,G. (2003) Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. Genome Res., 13, 37–45.
Proost,S. et al. (2012) i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. Nucleic Acids Res., 40, e11.
Renwick,J.H. (1971) The mapping of human chromosomes. Annu. Rev. Genet., 5, 81–120.
Robert,N.S.M. et al. (2022) Emergence of distinct syntenic density regimes is associated with early metazoan genomic transitions. BMC Genomics, 23, 143.
Rodelsperger,C. and Dieterich,C. (2010) CYNTENATOR: progressive gene order alignment of 17 vertebrate genomes. PLoS One, 5, e8861.
Simakov,O. et al. (2022) Deeply conserved synteny and the evolution of metazoan chromosomes. Sci. Adv., 8, eabi5884.
Simakov,O. et al. (2013) Insights into bilaterian evolution from three spiralian genomes. Nature, 493, 526–531.
Sun,M. and Zhang,J. (2019) Chromosome-wide co-fluctuation of stochastic gene expression in mammalian cells. PLoS Genet., 15, e1008389.
Veron,A.S. et al. (2011) Close 3D proximity of evolutionary breakpoints argues for the notion of spatial synteny. BMC Genomics, 12, 303.
Wang,Y. et al. (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res., 40, e49.
Winter,S. et al. (2016) Finding approximate gene clusters with gecko 3. Nucleic Acids Res., gkw843.
Zimmermann,B. et al. (2019) Ancient animal genome architecture reflects cell type identities. Nat. Ecol. Evol., 3, 1289–1293.